

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_nzp model

Run on: August 28, 2004, 07:42:36 ; Search time 207.5 Seconds
(without alignments)
5233.801 Million cell updates/sec

Title: US-10-676-079-3
Perfect score: 3119
Sequence: 1 ctgagcttcctgactctccg.....atactagctctgactcgt 1721

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame_nzp.model -DEV=xlp
-O=/cgn2_1/USPTO_spool_p/US10676079/runat_28082004_084225_23813/app_query.fasta_1.1863
-DB=SPTRMBL_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosun62 -TRANS=human40.cdi -LIST=45
-LOCAL=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10676079_@CGN_1_1_293_@runat_28082004_084225_23813 -NCPU=6 -ICPU=3
-NO_MAP -LARGEBUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLONG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	2838	91.0	543	4 Q9Y251	Q9Y251 homo sapien

2	2817	90.3	545	4 Q9UL39	Q9UL39 homo sapien
3	2282	73.2	545	6 Q9MYO0	Q9MYO0 bos taurus
4	2150	68.9	535	11 Q8K3K3	Q8K3K3 mus musculus
5	2123	68.1	536	11 Q9QZP8	Q9QZP8 rattus norv
6	1645.5	52.8	523	13 Q90YK5	Q90YK5 gallus gall
7	1154.5	37.0	592	4 Q9HB37	Q9HB37 homo sapien
8	1146.5	36.8	592	4 Q8WMQ2	Q8WMQ2 homo sapien
9	1015.5	33.6	548	4 Q8WMQ1	Q8WMQ1 homo sapien
10	936.5	30.0	534	4 Q9HB38	Q9HB38 homo sapien
11	897.5	28.8	480	4 Q9HB39	Q9HB39 homo sapien
12	696	22.3	515	5 Q8T108	Q8T108 bombyx mori
13	416	13.3	521	10 Q9SDA1	Q9SDA1 arabidopsis
14	416	13.3	543	10 Q9FF10	Q9FF10 arabidopsis
15	406.5	13.0	559	16 Q89FP9	Q89FP9 bradyrhizob
16	392.5	12.6	544	10 Q8H615	Q8H615 oryza sativ
17	382	12.2	527	10 Q9LRC8	Q9LRC8 scutellaria
18	363	11.6	536	10 Q9FZP1	Q9FZP1 arabidopsis
19	352.5	11.3	516	10 Q9FLK8	Q9FLK8 arabidopsis
20	352.5	11.3	539	10 Q8L608	Q8L608 arabidopsis
21	169.5	5.4	190	10 Q82604	Q82604 arabidopsis
22	160	5.1	935	5 Q9VE79	Q9VE79 drosophila
23	135	4.3	411	16 P72895	P72895 synecocyst
24	130.5	4.2	493	17 Q9HK01	Q9HK01 thermoplasma
25	126	4.0	408	3 Q9HEZ1	Q9HEZ1 phanerochaete
26	126	4.0	408	3 Q9HEZ2	Q9HEZ2 phanerochaete
27	124	4.0	366	10 Q8GSP4	Q8GSP4 lotus japon
28	118.5	3.8	2319	3 Q96U00	Q96U00 neurospora
29	117	3.8	489	10 Q9SS90	Q9SS90 arabidopsis
30	116.5	3.7	617	12 Q40936	Q40936 measles vir
31	114.5	3.7	1169	5 Q869K5	Q869K5 dictyostell
32	114	3.7	398	16 Q8FA10	Q8FA10 leptospira
33	114	3.8	832	4 Q75427	Q75427 homo sapien
34	114	3.7	1449	10 Q7XPA4	Q7XPA4 oryza sativ
35	113.5	3.6	287	2 Q59144	Q59144 aeromonas b
36	113.5	3.6	617	12 Q83295	Q83295 measles vir
37	113.5	3.6	1829	2 Q9RH44	Q9RH44 pantoea agg
38	113	3.6	390	17 Q8TPH7	Q8TPH7 methanosaarc
39	113	3.6	641	17 Q97B86	Q97B86 thermoplasma
40	112.5	3.6	617	12 Q83647	Q83647 measles vir
41	112.5	3.7	692	16 Q81ZV3	Q81ZV3 streptomyces
42	112.5	3.6	1276	16 Q82J10	Q82J10 streptomyces
43	112	3.6	636	16 Q88RX6	Q88RX6 lactobacilli
44	112	3.7	680	5 Q95TU3	Q95TU3 drosophila
45	112	3.7	1224	5 Q9VXW2	Q9VXW2 drosophila

ALIGNMENTS

RESULT 1
ID Q9Y251 PRELIMINARY; PRT; 543 AA.
AC Q9Y251,
DT 01-NOV-1999 (TRMBLrel. 12, Created)
DT 01-NOV-1999 (TRMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TRMBLrel. 25, Last annotation update)
DE HEPARANASE (HPSB protein).
GN HPA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OC NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=95321249; PubMed=10395326;
RA Hulett M.D., Freeman C., Hamdorf B.J., Baker R.T., Harris M.J.,
Parish C.R.;
RT "Cloning of mammalian heparanase, an important enzyme in tumor
invasion and metastasis,"
RL Nat. Med. 5:803-809(1999).
RN (2)
RP SEQUENCE FROM N.A.
RA Vlodevsky I., Friedman Y., Elkin M., Alingorn H., Atzmon R.,

RA Ishai-Michaeli R., Bitan M., Pappo O., Peretz T., Michal I.,
RA Spector L., Becker I.;
RT "Mammalian heparanase: a novel gene involved in tumor progression and
RT metastasis";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99377052; PubMed=10446189;
RA Toyoshima M., Nakajima M.;
RT "Human heparanase. Purification, characterization, cloning, and
RT expression";
RL J. Biol. Chem. 274:24153-24160(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=9935379; PubMed=10405343;
RA Kusie P.H., Hulmes J.D., Ludwig D., Patel S., Navarro E.C.,
RA Soddon A.P., Giorgio N.A., Bohlen P., Patel S., Navarro E.C.,
RT "Cloning and Functional Expression of a Human Heparanase Gene";
RL Biochem. Biophys. Res. Commun. 261:183-187(1999).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marisla K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshitoki S., Carinci P., Prange C.,
RA Raha S.S., Loquiano N.A., Peters G.J., Abramson R.D., Mulhaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyanski M.I., Skaleja U., Smallie D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strauberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RX MEDLINE=1615154; AAD45379.1; -
DR EMBL/ AFI44325; AAD41342.1; -
DR EMBL/ AFI55510; AAD54941.1; -
DR EMBL/ AFI52376; AAD54669.1; -
DR EMBL/ BC051321; AAH51321.1; -
DR Genev. HGNC:5164; HPSE.
DR GO; GO:0004566; F:beta-glucuronidase activity; TAS.
DR GO; GO:0007125; P:invasive growth; TAS.
DR GO; GO:0006029; P:proteoglycan metabolism; TAS.
DR InterPro; IPR005199; Glyco_hydro_79N.
DR Pfam; PF03662; Glyco_hydro_79n; I.
SQ SEQUENCE 543 AA; 61176 MW; AD262EC267334AB2 CRC64;

Alignment Scores:

Prod. No.: 6.3e-222 Length: 543
Score: 2838.00 Matches: 542
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.82% Mismatches: 0
Query Match: 90.99% Indels: 0
DB: 4 Gaps: 0

US-10-676-079-3 (1-1721) x Q9Y251 (1-543)

QY 63 ATGCTGCTGCGCTCGAAGCCTGCGCGCGCGCGCTGATGCTGCTCTGCGGCGG 122

Db
1 MetLeuLeuArgSerIysProAlaLeuProProLeuMetLeuLeuLeuGlyPro 20
QY CTGGATCCCCCTCCCTGGCGCCGCGCCGACCTCGCAAGCACAGGACGCGGAC 182
Db
21 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaMetValAlaSer 40
QY CTGGACCTTCTTCACCCAGAGCGCGCTGCACCTGGTAGGCCCTCGTCTGTCCTGAC 242
Db
41 LeuAspPhePheThrGlnGlnProLeuH1sLeuValIsrProSerPheLeuSerValThr 60
QY ATTGAAGCCCAACCTGGCCACGAGCCCGCGGTTCTCTATCTCTGGGTTCTCCAAAGCTT 302
Db
61 IleAspAlaAsnLeuAlaThrAspProArgPheLeuIleLeuLeuIlySerProIlysLeu 80
QY CGTACCTTGGCCGAGAGGCTTGCTCTCGGCTACCTGAGGTTGGTGGCCCAAGACAGAC 362
Db
81 ArgThrLeuAlaArgIlyLeuSerProAlaIlyLeuArgPheGlyIlyThrIysThrAsp 100
QY TTGCTTAATTTTCGATCCCAAGAGATCAACCTTGAAGAGAAATTACTGGCAATCT 422
Db
101 PheLeuIlePheAspProIlysPheIlySerThrPheGlyGlyIlyAspIlyTrpGlnSer 120
QY CAAGTCACCAAGATTTTGGCAAAATATGATGCATCCCTCTGATGTGGAGAGAAAGTAA 482
Db
121 GlnValAsnGlnAspIleCysIlyIlySerIleProProAspValGlnGlyIlysLeu 140
QY CGGTTGAATGCGCCCTACACAGAGCAATGCTACTCCGAGAACACACACAGAAAGTTTC 542
Db
141 ArgLeuGlnIlyProIlyGlnGlnGlnLeuLeuArgGlnIlyIlyGlnIlysPhe 160
QY 543 AAGAACAGCACTACTCAAGAAAGCTGTGATGTGCTATACACTTTGCAAACTGCTCA 602
Db
161 LysAsnSerThrIlySerIlySerSerValAspValLeuIlyThrPheAlaAsnIlySer 180
QY GGACTGACTGATCTTTGGCCCTTAATGCGTTATTAAGAACAGCAATTTGCACTGAGAAC 662
Db
181 GlyLeuAspLeuIlePheGlyLeuAlaLeuLeuArgThrAlaAspLeuGlnIlyPhe 200
QY AGTTTAATGCTGATGTTGCTCTGACACTGCTCTTCCAAAGGAGTAAACATTTCTGG 722
Db
201 SerSerAsnAlaGlnLeuLeuLeuLeuAspIlyIlySerSerIlyGlyIlyIlySerIly 220
QY 723 GAACTAGGCAATGAACCTTAACAGTTTCTTAAGAAAGCTGTATTTTCATCAATGGGTCG 782
Db
221 GlnLeuGlyAsnGlnProAsnSerPheLeuIlyValAlaAspIlePheIleAsnIlySer 240
QY CAGTTAGGAGAAATATATTTCAATTGCAATTAACCTTAAGAAAGCTTCAAAAT 842
Db
241 GlnLeuGlyGlnAspPheIleGlnLeuH1sIlyLeuArgIlySerThrPheIlyAsn 260
QY GCAAACTGATGCTGCTGATGTTGCTGACCTCGAAGAAAGAGGCTAAAGATGCTGAAG 902
Db
261 AlaIlyLeuIlyIlyIlyProAspValGlyGlnProArgIlyIlyThrAlaIlysLeuIly 280
QY AGCTTCTTAAGAGCTGTGAGAAAGTATGATTGCTTACATGATGATCACTACTATTTG 962
Db
281 SerPheLeuIlyLeuIlyGlyGlyValIleAspSerValThrIlyIlyIlyIlyIly 300
QY AATGAGACGAGCTGCTACCGAGGAAGATTTTCAACCGATGATTAATGACATTTTAT 1022
Db
301 AsnGlyArgThrAlaThrArgGlnAspPheLeuAsnProAspValLeuAspIlePheIle 320
QY TCATGCTGCAAAAAGTTTCCAGTGTGAGAGACACGAGCTTGGCAAGAGCTGCG 1082
Db
321 SerSerValGlnIlyValPheGlnValIlyGlnSerThrArgProGlyIlyIlyValIly 340
QY TTAGAGAAACAAGCTCTGCATATGAGAGCGAGCGCTTGTCTATCGACACTTTGCA 1142
Db
341 LeuGlyGlyIlyThrSerSerAlaIlyIlyGlyIlyAlaProLeuLeuSerAspThrPheAla 360
QY GCTGCTTATATGCTGATTAATGAGGCTGTGACGCCGGAATGGGAATGAAGTGTG 1202


```

Db      361 PheAlaIaGlyPheMetTrpLeuAspLysLeuGlyLeuSerXlaaGmetGlyIleGlu 380
QY      1197 GTGTGATGAGCAAGATATCTTTGGAGCAGAAACCACTATTAGATGAGAAACTTC 1256
Db      381 ValValMetArgGlnValIlePhePheGlyAlaGlyAsnThrIleuValAspGluAspPhe 400
QY      1257 GATCTTACCTGATATTTGGCTATCTCTTCTGTCAAGAAATTTGGTGGCACCAGGTG 1316
Db      401 AspProLeuProAspTyrTrpLeuSerLeuLeuPheLysLeuValGlyThrVal 420
QY      1317 TTAAATGGCAAGCGTGCAGAGTTCAAAGAGAAAGCTTGAGATACCTTCATGTGACA 1376
Db      421 LeuMetAlaSerValGlnGlySerLysArgArgLysLeuArgValTyrLeuHisGlyThr 440
QY      1377 AACACTGACATCCAAAGTATTAAGAAGAGATTAATCTGTATGCGCATTAACCTCAT 1436
Db      441 AsnThrAspAsnProAlaGlyTyrLysGlnGlyAspLeuThrLeuTyrAlaIleAsnLeuHis 460
QY      1437 AACGTCACCAAGTACTTGGCTTACCTCATCTTTTCTAACAGCAAGTGATTAATAC 1496
Db      461 AsnValThrLysThrLysLeuArgLeuProTyrProPheSerAsnLysGlnValAspLysTyr 480
QY      1497 CTCTTAAGACCTTTGGGACCTCATGTGATTACTTTCCAAATCTGTCCAACTCATGTCTA 1556
Db      481 LeuLeuArgProLeuGlyProHisGlyLeuLeuSerLysSerValGlnLeuAsnGlyLeu 500
QY      1557 ACTCTAAAGATGGTGATGATCAAAACCTTGCAACCTTAAATGAAAAACCTCCGGGCA 1616
Db      501 ThrLeuLysMetValAspAspGlnThrLeuProLeuMetGlnLysProLeuAsnArgPro 520
QY      1617 GGAAGTTCACTGGGCTTGCCAGCTTTCTCATATATGTTTTTGTGATTAAGAAATGCCAA 1676
Db      521 GlySerSerLeuGlyLeuProAlaPheSerTyrSerPhePheValIleArgAsnAlaLys 540
QY      1677 GTTGCCTGTGCATC 1691
Db      541 ValAlaIaIaCysIle 545

```

RESULT 3

```

Q9MYT0 ID Q9MYT0 PRELIMINARY; PRT; 545 AA.
AC      09MYT0;
DT      01-OCT-2000 (TEMBLrel. 15, Created)
DT      01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT      01-JUN-2002 (TEMBLrel. 21, Last annotation update)
OS      Heparanease.
DS      Bos taurus (Bovine).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC      Bovidae; Bovinae; Bos.
OX      NCBI_TaxID=9913;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Placenta;
RA      Kizaki K., Nakano H., Takahashi T., Imai K., Hashizume K.;
RT      "Expression of Heparanase mRNA in Bovine Placenta During Gestation.";
RL      Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AF81160; AAF87301.2; -
DR      InterPro: IPR005199; Glyco_hydro_79N.
DR      Pfam: PF03662; Glyco_hydro_79n; I.
SQ      SEQUENCE 545 AA; 61076 MW; FAC4BDFD85B933 CRC64;

```

Alignment Scores:

```

Pred. No.:      9,98e-177      length:      545
Score:          2282.00      Matches:      435
Percent Similarity:      86.24%      Conservative: 35
Best Local Similarity:      79.82%      Mismatches: 73
Query Match:      73.16%      Indels:      2
DB:              6              Gaps:      1

```

US-10-676-079-3 (1-1721) x Q9MYT0 (1-545)

```

QY      63 ATCTGCTGGCGCTCGAAGCTGCGCTGCGCGCGCTGATGCTGTG-----CTCTTG 116
Db      1 MetLeuAlaCysArgLysProGlyLeuArgProProLeuLeuLeuLeuProLeuLeu 20
QY      117 GGGCGCGCTGGTCCCTCTCCCTCGCGCGCTCCCGGACCTGGCGAACAAGCAAGCGTC 176
Db      21 GlyProLeuGlyProCysSerProGlyThrProAlaAlaAlaAlaProAlaAspAla 40
QY      177 GTGAGCTGGACTTTCTTCAACCCAGAGCCGCTGCACTGGTGAAGCCCTGTCTCTGCC 236
Db      41 AlaGluLeuGlnPhePheThrGluArgProLeuHisLeuValSerProAlaPheLeuSer 60
QY      237 GTACCAATTGAGCCCAACCTGGCCAGACCGCGGTTCTCATCTCTCGGTTCTCA 296
Db      61 PheThrIleAspAlaAsnLeuAlaThrAspProArgPhePheThrPheLeuGlySerSer 80
QY      297 AACCTTCGATCCTTGGCGCAGAGCTGTGCTCCCTGCGCTGATGAGCTTGGTGGCACAAG 356
Db      81 LysLeuArgThrLeuAlaArgGlyLeuAlaProAlaTyrLeuArgPheGlyGlyAsnLys 100
QY      357 ACAGACTTCCTAATTTTCGATCCCAAGAAAGAAATCAACCTTTGAAAGAGAAATTAAG 416
Db      101 GlyAspPheLeuIlePheAspProLysLysGlnProAlaPheGlnGlnArgSerTyrTrp 120
QY      417 CAATTCGAATCAACAAGATATTGCAAAATATGATTCATCCCTCTGATGTGAGAG 476
Db      121 LeuSerGlnSerAsnGlnAspIleCysLysSerGlySerIleProSerAspValGlnGlu 140
QY      477 AAGTTACGGTGGAAAGGCGCTTACAGAGCAATTGCTCTCCGAGAACACTACAGAA 536
Db      141 LysLeuArgLeuGlnTrpProPheGlnGlnValLeuLeuArgGlnGlnTyrGlnLys 160
QY      537 AAGTTCAAGAACAGCACTACTCAAGAGCTGTGATGTGCTATACATTTTGGCAAC 596
Db      161 LysPheThrAsnSerThrLysSerArgSerSerValAspMetLeuThrThrPheAlaSer 180
QY      597 TGCTCAAGACTGGAATCTTTGGCTTAATGCTTATTAAGAACAGCAATTTGCAAG 656
Db      181 CysSerGlyLeuAsnLeuIlePheGlyValAsnAlaLeuLeuArgThrThrPheMetHis 200
QY      657 TGGACAGTTCTAATGCTCAAGTGTCTCCGAGCTACTGCTCTCAAGGGGTTAACATT 716
Db      201 TrpAspSerSerAsnAlaGlnLeuLeuAspTyrCysSerSerLysAsnTyrAsnIle 220
QY      717 TCTTGGAACTAGGCAATGAACCTTAACAGTTTCTTAAAGAGCTGATATTTTCATCAAT 776
Db      221 SerTrpGlnLeuGlyAsnGlnProAsnSerPheGlnArgLysAlaGlyIlePheIleAsn 240
QY      777 GGGTCGCACTTGAAGAAAGTATTAATTCATATTCATTAACCTTAAAGAAAGTCCACTTC 836
Db      241 GlyArgGlnLeuGlyGlnAspPheIleGlnPheArgLysLeuLeuGlyLysSerAlaPhe 260
QY      837 AAAAATGCAAAACTCATGCTCTGATGTTGGTGCAGCTCGAAGAAAGCGGTAAAGTG 896
Db      261 LysAsnAlaLysLeuTyrGlyProAspIleGlyGlnProArgArgAsnThrValLysMet 280
QY      897 CTGAAGAGCTCTCTGAAGCGTGTGAGAGAGTATGATTCATGATCAATGCAATCACTAC 956
Db      281 LeuLysSerPheLeuLysAlaGlyGlyValIleAspSerValThrTrpHisnIleTyr 300
QY      957 TATTGAATGACAGCACTGCTACAGAGAAAGATTTTCTAAACCTTGATGATTTGACATT 1016
Db      301 TyrValAsnGlyArgGlyLeaIleThrLysGlnAspPheLeuAsnProAspIleLeuAspThr 320
QY      1017 TTATATTCACTGTGAAAGAAAGTTTCCAGTGCGTGAAGACACCGGCTGGCAAGAG 1076
Db      321 PheIleSerSerValGlnLysThrLeuArgIleValGlnLysIleArgProLeuLysLys 340
QY      1077 GTCTGTTAGAGAAACAAGCTCTGATATGAGAGCGAGCGGCTTGTGCTATTCGACACC 1136
Db      341 ValTrpLeuGlyGlnThrSerSerAlaPheGlyGlyAlaProPheLeuSerAsnThr 360
QY      1137 TTTCAGCTGGCTTTATGTGGCTGATTAATTTGGGCTGTCAAGCCGCAATGGAAATAGAA 1196

```



```

Db      361 PheAlaAlaGlyPheMetTrpLeuSerPheGlyLeuSerAlaArgGlyIleGlu 380
Qy      1197 GGGGAGATGAGCAGATATCTTTGAGAGAGAAATACCATTTAGTGATGAAAACTTC 1256
Db      381 ValValMetArgGlnValLeuPheGlyAlaGlySerTrpHisLeuValAspGlyAspPhe 400
Qy      1257 GATCCCTTACCTGATATATGGGATCTCTTCTGTTCAAGAAATGGTGGCCAAAGGCG 1316
Db      401 GtUProLeuProAspTrpTrpLeuSerLeuPheValLeuValAlaGlyAsnValVal 420
Qy      1317 TTAATGGCAAGCGGTGCAAGGTTCAAGAGAAAGCTTCAAGTATACCTTCAATGGCACA 1376
Db      421 LeuMetAlaSerValIleGlyProAspArgSerLysPheArgValTrpLeuHisCysThr 440
Qy      1377 AACACTGACAAATCCAAAGGTATTAAGAGAGATTTACTCTGTATGCCATAAACCCTCCAT 1436
Db      441 AantTrpLysHisProArgTrpLysGlyAspLeuThrLeuValAlaLeuAsnLeuHis 460
Qy      1437 AACGTCAACCAAGTACTTGGCGTTACCTATCCCTTTTCTTCAACAGCAAGTGATAATAC 1496
Db      461 AnValThrLysHisLeuGluLeuProHisHisLeuPheAsnLysGlnValAspLysTrp 480
Qy      1497 CTTTCAAGACTTTGGGACCTCATGATTAATCTTCCAAATCTGCCAATCAATGCTCTTA 1556
Db      481 LeuIleLysProSerGlyThrAspGlyLeuLeuSerLysValGlnLeuAsnGlyGln 500
Qy      1557 ACTCTTAAGATGTGATGATCAAACTTGCCCACTTTAATGAAAAAACCCTCCGCGCA 1616
Db      501 IleLeuLysMetValAspGlnGlnThrLeuProAlaLeuThrGluLysProLeuHisPro 520
Qy      1617 GGAAGTTCACTGGGCTTGCAAGCTTCTCATATAGTTTGTGTGATAAGAAATGCCAA 1676
Db      521 GlySerSerLeuGlyMetProProPheSerTrpGlyPhePheValIleArgAsnAlaLys 540
Qy      1677 GTTGCTGCTTGCAATC 1691
Db      541 ValAlaAlaCysIle 545

RESULT 4
ID      08KXK3      PRELIMINARY;      PRT;      535 AA.
AC      08KXK3;
DT      01-OCT-2002 (TREMblrel. 22, Created)
DT      01-OCT-2002 (TREMblrel. 22, Last sequence update)
DT      01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE      Heparanase.
GN      HPSE OR HPA.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN=FVB;
RX      MEDLINE=2354683; PubMed=12466851;
RA      The PANTOM Consortium.
RA      the RIKEN Genome Exploration Research Group Phase I & II Team;
RT      "Analysis of the mouse transcriptome based on functional annotation of
RT      60,770 full-length cDNAs."
RL      Nature 420:563-573 (2002).
DR      EMBL; AY077467; BAB76083.1; -.
DR      EMBL; AK040471; BAC30600.1; -.
DR      MGD; MGI:1343124; Hspe.
DR      InterPro; IPR005199; Glyco_hydro_79N.
DR      Pfam; PF03662; Glyco_hydro_79n; 1.

```

```

SQ      SEQUENCE      535 AA;      60065 MW;      6E73A8302FBA0DF CRC64;
Alignment Scores:
Pred. No.:      5,33e-166      Length:      535
Score:      2150.00      Matches:      407
Percent Similarity:      86.25%      Conservative:      51
Best Local Similarity:      76.65%      Mismatches:      73
Query Match:      68.93%      Indels:      0
DB:      11      Gaps:      0

US-10-676-079-3 (1-1721) x 08KXK3 (1-535)
Qy      99 CTGATGCTGCTGCTGCTGCGGCGGTGATCCCTCTCCCGTGGCCGCTCCGACT 158
Db      5 LeuLeuLeuTrpLeuTrpGlyProLeuGlyAlaLeuAlaGlnGlyAlaProAlaGlyThr 24
Qy      159 GCGCAAGCAGCAGACGCTGTGGAACCTTGACCTTTCACCCAGAGACCGCTGCACTGGTG 218
Db      25 AlaProThrAspAspValValAspLeuGluPheTrpLysArgProLeuArgSerVal 44
Qy      219 AGCCCTGCTGCTGCTGCTCCGTCACATTTGACCGCCAGCAGGACCGGCGGTTCTTC 278
Db      45 SerProSerPheLeuSerIleThrIleAspAlaSerLeuAlaThrAspProArgPheLeu 64
Qy      279 ATCTCTGCGGTTCTTCAAGCTTGCTGTAACCTTGCCAGAGGCTTGTCTGCTGTAACCTG 338
Db      65 ThrPheLeuGlySerProArgLeuArgAlaLeuAlaArgGlyLeuSerProAlaTrpLeu 84
Qy      339 AGCTTGGTGGCAGCAGACAGACTTCTTAATTTTGCATCCCAAGAGAAATCAACTTT 398
Db      85 ArgPheGlyGlyThrLysThrAspPheLeuIlePheAspProAspLysGluProThrSer 104
Qy      399 GAAAGAGAGAGTACTGCGAATCTCAAGTCCAGACCAAGATTTGGAATATGATCCATC 458
Db      105 GlnGluArgSerTrpLysSerGlnValAsnHisAspIleCysArgSerGluProVal 124
Qy      459 CCTCTGATGTGAGAGAGATTAAGTTCAGTGAATGAGCCCTACAGAGCAATGTGACTC 518
Db      125 SerAlaAlaValLeuArgLysLeuGlnValGluTrpProPheGlnGluLeuLeuLeu 144
Qy      519 CGAAGAACACTACCAAGAAAAGTTCAAGAACAGCACTTACTCAAGAACTCTGATAGTGTG 578
Db      145 ArgGlnGlnTrpGlnLysGlnLysLeuPheLysAsnSerThrTrpSerArgSerValAspMet 164
Qy      579 CTAATACCTTTGCAAACTGCTCAGACGTCGACCTTGATCTTGGCTAAATGCGTATTA 638
Db      165 LeuTrpSerPheAlaLysCysSerGlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeu 184
Qy      639 AGAAGACAGATTTGCAAGTGAACAGTTCTAATGCTCAGTTGCTCCTGGACTACTGCTCT 698
Db      185 ArgThrProAspLeuArgTrpAsnSerSerAsnAlaGlnLeuLeuLeuAspTrpCysSer 204
Qy      699 TCCAAAGGGATATTAACATTTCTTGGAACTAGGCAATGAACCTTAACAGTTTCTTAAGAG 758
Db      205 SerLysGlyTrpAsnIleSerTrpGluLeuGlyAsnGluProAsnSerPheTrpLysLys 224
Qy      759 GGTGATATTTTCAATGAAGGCGTCGACGTAAGAGAAATTAATTCATTTGATGATAACT 818
Db      225 AlaHisIleLeuIleAspGlyLeuGlnLeuGlyGluAspPheValGluLeuHisLysLeu 244
Qy      819 CTAAGAAAGTCCACCTTCAAAATGCAAAATCTTAATGCTCAGTTGCTGGAATGATTAATCA 878
Db      245 LeuGlnArgSerAlaPheGlnAsnAlaLysLeuTrpGlyProAspIleGlyGlnProArg 264
Qy      879 AGAAAAGCGGCTAAGAGCTGAAGAGCTTCTGAAGGCTGTGAGAAATGATTAATCA 938
Db      265 GlyLysThrValIleLeuLeuArgSerPheLeuLysAlaGlyGlyGluValIleAspSer 284
Qy      939 GTTACATGGCAATCACTATTTGAATGAAGCAGCATCTGCACAGAGAAATTTTAAAC 998
Db      285 LeuThrTrpHisIleTrpLysLeuAsnGlyArgIleAlaThrLysGlnAspPheLeuSer 304
Qy      999 CCGATGTATGAGACATTTTATTCATCTGTGCAAAAAGTTTTCAGAGTGGTTGAGAGC 1058

```

```

Db      305 SerAspValLeuSprThrPheIleLeuSerValGlnIlySleuLysValThrIlySglu 324
Qy      1059 ACCAGGCGCTGGCAAGAGGTCTGGTAGAGAAACAAGCTTCGATTTGGAGCGGAGCG 1118
Db      325 IleThrProGlyLysLysValThrPheGlyGlnIlySerSerAlaIyrgIyGlyAla 344
Qy      1119 CCCTTGTATCCGACACCTTTCAGCTGGCTTATATGTGGCTGATTAATTTGGGCTGTCA 1178
Db      345 ProLeuLeuSerAenThrPheAlaIagIyPheMetThrPheAepLysLeuGlyLeuSer 364
Qy      1179 GCGCGAATGGGAATAGAAAGTGGTATAGAGCAATATCTTTGGAGCAAGAACTTACAT 1238
Db      365 AlagIlyMetGlyIleGlyValAlaIleCArgGlnValPhePheGlyAlaGlyAenIyThIs 384
Qy      1239 TTATGGATGAAAACTTCGATTCCTTACCTGATTAATTTGGCTATCTCTTTCAGAAA 1298
Db      385 LeuValAlaSerLysLysPheGlyProLeuProAepIyrrPheLeuSerLeuPheLysLys 404
Qy      1299 TTGGTGGGACCAAGGTATTATGGCAAGCGTGCAGCTTCMAAGAGAAAGCTTGA 1358
Db      405 LeuValGlyProAenValLeuLeuSerArgValIlyGlyProAepArgSerLysLeuArg 424
Qy      1359 GTATACCTTCAATTCGCAAAACACGACAAATCCAAAGGTATTAAAGAGATTTAACTGTG 1418
Db      425 ValTyrLeuIlyIleCysThrAenValTyrIlyIleProAenIyrrGlnIlyAepLysLeu 444
Qy      1419 TATGCACTAAACCTCCATAACGTCACCAAGTACTTGGGCTTACCCCTATCTTTCTTAC 1478
Db      445 TyrValLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 464
Qy      1479 AACCAAGTGAATTAATACCTTTTGAACCTTTGGACCTGATGATTAATTTCCAAATCT 1538
Db      465 LysProValAepThrTyrLeuLeuLysProSerGlyProAepGlyLeuLeuSerLysSer 484
Qy      1539 GTCCAACTCAATGCTCTCACTCTTAAAGTGTGATGATGATGATGATGATGATGATGATG 1598
Db      485 ValGlnLeuLysGlnIlyIleLeuLysMetValAepLysLysLysLysLysLysLys 504
Qy      1599 GAAAAACCTTCGCGGACCAAGGTTCACTGGGCTTGGCAGCTTCTCATATAGTTTGT 1658
Db      505 GlnLysProLeuProAenIyrrGlySerAlaLeuSerLeuProAenIyrrGlyPhePhe 524
Qy      1659 GTGATTAAGAAATGCCAAAGTGTGCTTGCATC 1691
Db      525 ValIleArgAenAlaIySleAlaIyCysIle 535

RESULT 5
Q9QZF8      PRELIMINARY;      PRT;      536 AA.
ID      Q9QZF8;
AC      Q9QZF8;
DT      01-MAY-2000 (TREMBLrel. 13, Created)
DT      01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      Heparanase.
GN      HEP.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Podiyana K.A., Yokote H., Sakaguchi K., Ikuta M., Yanagishita M.;
RT      "Heparanase from parathyroid cell line.";
RL      Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF184967; AAF04563.1; -
DR      InterPro; IPR005199; Glyco_hydro_79n.
DR      Pfam; PF03662; Glyco_hydro_79n; 1.
SQ      SEQUENCE 536 AA; 60569 MW; 6208B1FD9E28421 CRC64;

```

Alignment Scores: 8.36e-164 Length: 536
 Pred. No.: 2123.00 Matches: 405

```

Percent Similarity: 85.23% Conservative: 51
Best Local Similarity: 75.70% Mismatches: 79
Query Match: 68.07% Indels: 0
DB: 11 Gaps: 0

US-10-676-079-3 (1-1721) x Q9QZF8 (1-536)

Qy      87 CTGCGCGCGCGCTGATGCTGTGCTCTGCGGCGCGCTGCGCTCCCTGCGCGCGC 146
Db      2 LeuArgProLeuLeuThrPheLeuTrpGlyArgLeuArgAlaLeuThrGlnIlyThr 21
Qy      147 CTGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 206
Db      22 ProAlaGlyThrAlaProThrLysAepValAlaAepLysGlnPheIyrrThrySAArgLeu 41
Qy      207 CTGACCTGCGTGGAGCCCTCGTTCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 266
Db      42 PheGlnSerValSerProSerPheLeuSerIleThrIleAepAlaSerLeuAlaThrAep 61
Qy      267 CCGCGGTTCTCATCTCTCTGCGGTTCTCGAAAGCTTCGATCTTGGCCAGAGGCTTGTCT 326
Db      62 ProArgPheLeuThrPheLeuGlySerProArgLeuArgAlaLeuAlaArgIlyLeuSer 81
Qy      327 CCGCGGTTCTGAGGTTGGTGGCACCAAGACGATCTTCAATTTTGCATCCGAAAG 386
Db      82 ProAlaTyrLeuArgPheGlyGlyThrLysThrAepPheLeuIlePheAepProAenLys 101
Qy      387 GAATCAACCTTTGAAAGAGAGAGTTCAGTCAATCTCAAGTCAACGAGATATTTCGAAA 446
Db      102 GlnProThrSerGlnLysSerIyrrTrpGlnSerGlnAepAenAepIleCysGly 121
Qy      447 TATGATTCATCCCTCTGATGTGAGAGAGATTAAGGTGAGATGCGCTTGAACGAGAG 506
Db      122 SerGlnArgValSerAlaAepValLeuArgLysLeuGlnMetGlnTrpProPheGlnIly 141
Qy      507 CAATTGCTCTCGGAAACACATCACGAAAGAGTTCAAAAGCGACCTTCAAGAAAGC 566
Db      142 LeuLeuLeuLeuArgGlnIlyTrpGlnArgGlnPheLysAenSerThrIyrrSerArgSer 161
Qy      567 TCTGTAGATGTGATTAACATTTTGGCAACCTGCTCAGACCTGGAATTTGATCTTTGGCCTA 626
Db      162 SerValAepMetLeuIyrrSerPheAlaIyrrCysSerArgLeuAepLeuIlePheGlyLeu 181
Qy      627 AATGCGTTATTAAAGACAGCAAGTTTGCAGTGAACAGTTTCAATGCTCAGTTGCTCTG 686
Db      182 AenAlaLeuLeuArgThrProAepLeuArgTrpAenSerSerAenAlaGlnLeuLeu 201
Qy      687 GACTACTGCTCTTCCAAAGGCTTAACTTTTGGGAACTTGGGAACTGGAATTAACCTAACAT 746
Db      202 AenTyrCysSerSerIyrrGlyTrpAenIleSerTrpGlnLeuGlyAenGlnProAenSer 221
Qy      747 TTCTTAAGAAAGGCTGATATTATTTTCATCAATGGGTCGACATTAAGAGAAATTAATTC 806
Db      222 PheTrpLysLysAlaGlnIleSerIleAepGlyLeuGlnLeuGlyLysPheValGln 241
Qy      807 TTGATTAACCTTAAAGAAAGTCCACCTTCMAAATGCAAACTATAGTCTGATGTT 866
Db      242 LeuIleLysLeuLeuLysSerAlaPheGlnAenAlaIyLeuLysIyrrProAepIle 261
Qy      867 GGTCAAGCTTCGAAAGAGAGCGCTAAGCTGCAAGAGCTTCTGAAGGCTGTGGAGAA 926
Db      262 GlyGlnProArgIlyLysThrValLysLeuLeuArgSerPheLeuLysAlaGlyGlyGln 281
Qy      927 GTATGATTAATGATTCATGAGCTCACTCAATTTTAAAGAGAGAGCTCAACAGGAA 986
Db      282 ValIleAepSerLeuThrTrpIlyshIyrrTyrLeuAenGlyArgValAlaThrIlySglu 301
Qy      987 GATTTTCAACCCCTGATGATTTGACATTTTATTTTCACTGTGCAAAAGTTTTCAG 1046
Db      302 AspPheLeuSerSerAepValLeuAepThrPheIleLeuSerValGlnIlySleuLys 321
Qy      1047 GTGGTTGAGAGCACCGGCTGCGAAGAGGCTGTGTTAGAGAAACAAGCTTGCATAT 1106

```

```

Db      322 ValThrLysGluMetThrProGlyLysValTrpLeuGlyGlnThrSerSerAlaTyr 341
QY      1107 GGAGGCGGAGCGCCCTTGCTATCCGACCTTGAGCTGGCTTATGNGGCTGATAAA 1166
Db      342 GlyGlyGlyAlaProLeuLeuSerSerThrPheAlaAlaGlyPheMetTrpLeuSerLys 361
QY      1167 TTGGGCTGTGACCCCGAATGGGAATGAAAGTGGTATGAGGCAAGTATCTTTGGAGCA 1226
Db      362 LeuGlyLeuSerAlaGlnLeuGlyLeuValValMetArgGlnValPhePheGlyAla 381
QY      1227 GGAATCTACATTTAGTGGATGAAATCTTCGATCCCTTACCTGATTTAGTCTATCTT 1286
Db      382 GlyAlaThrLysLeuValAlaProGlnSerPheGluProLeuProAspLysTrpLeuSerLeu 401
QY      1287 CTGTTCAAGAAATTTGGTGGGACCAAGGTGTTAATGACCAAGGCTGGCAAGTTCAAGAGA 1346
Db      402 LeuPheLysLysLeuValGlyProLysValLeuMetSerArgValLysGlyProAspArg 421
QY      1347 AGAAGCTTCGAGTATACCTTCAATTCGACAAACACTGACATCCAAAGTATTAAGAGCA 1406
Db      422 SerLysLeuArgValLysLeuHisCysThrAsnValTyrHisProArgLysTrpArgGly 441
QY      1407 GATTTAATCTGTATGCCATAAACCCTCAATAAGCTCAACAAGTACTTGGCGTTACCTTAT 1466
Db      442 AspLeuThrLeuTyrValLeuAsnLeuHisAsnValThrLysHisLeuLysLeuProPro 461
QY      1467 CCTTTTCTAACAGAGAGTGAATTAATACCTTCTTAAGACCTTTGGGACCTGATGATTA 1526
Db      462 ProMetPheSerArgProValAspLysLysLysLeuLysProPheLysSerAspGlyLeu 481
QY      1527 CTTTCCAAATCTGTCCAACTCAATGCTTAACTCTTAAGATGGTGGATGATCAAACTTGG 1586
Db      482 LeuSerLysSerValGlnLeuAsnGlyGlnThrLeuLysMetValAspGlnGlnThrLeu 501
QY      1587 CCACTTTAATGAAAAACCTTCGCGGCAAGAGTTCACCTGCGCTTGGCAAGCTTTCTCA 1646
Db      502 ProAlaLeuThrGlnLysProLeuProAlaGlySerSerLeuSerValProAlaPheSer 521
QY      1647 TATAGTTTTTGTGATGAAGAAATGCAAGTTGCTGCTTGATC 1691
Db      522 TyrGlyPhePheValIleArgAsnAlaLysIleAlaAlaCysIle 536

RESULT 6
Q90YK5 PRELIMINARY; PRT: 523 AA.
ID      090YK5;
AC      090YK5;
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE      Heparanase.
OS      Gallus gallus (Chicken).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC      Gallus.
OX      NCBI_TaxId=9031;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      PubMed=11387326;
RA      Goldshmidt O., Zcharia E., Aingorn H., Gatica-Rangini Z., Atzmon R.,
RA      Michal I., Pecker I., Mitrani B., Vlodavsky I.;
RT      "Expression Pattern and Secretion of Human and Chicken Heparanase Are
RT      Determined by Their Signal Peptide Sequence.";
RL      J. Biol. Chem. 276:29178-29187(2001).
DR      EMBL; AY037007; AK82648.1; -.
DR      InterPro; IPR005199; Glyco_hydro_79N.
DR      Pfam; PF03662; Glyco_hydro_79n.1.
SQ      SEQUENCE 523 AA; 58386 MW; 8E80B7B18C9BF81 CRC64;

```

```

Query Match: 52.76% Indels: 11
DB: 13 Gaps: 3
us-10-676-079-3 (1-1721) x Q90YK5 (1-523)

QY      99 CTGATGCTGCTGCTCTGCGGCGCGTGGGCTCCCTCCCTGCGCCCTGCGCCGACT 158
Db      2 LeuValLeuLeuLeuValLeuLeuAlaValProPro----- 15
QY      159 GCGCAAGACAGGAGCTGTGACCTTGACCTTCCACCGAGGACCGCTGACCTGGTG 218
Db      16 -----ArgArgThrAlaGlnLeuGlnLeuGlyLeuAlaGlnProIleGlyAlaVal 32
QY      219 AGCCCTGCTGCTCTCCGCTCCGATGACGCGCACTGACGCGACCGCGGCTTCTC 278
Db      33 SerProAlaPheLeuSerLeuThrLeuAspAlaSerLeuAlaArgAspProArgPheVal 52
QY      279 ATCTCTGGGCTTCCAAAGCTTGATCTTGAGCAGAGGCTTGTCTCGCTGACTG 338
Db      53 AlaLeuLeuArgHisProLysLeuHisThrLeuAlaSerGlyLeuSerProGlyPheLeu 72
QY      339 AGCTTTGGTGGACCAAGACAGACTTCTTAATTTTCGATCCCAAGAAATCAACTTT 398
Db      73 ArgPheGlyGlyThrSerThrAspPheLeuIlePheAsnProAsnLysAspSerThrTrp 92
QY      399 GAAAGACAGTATTCTGCAATCTCAAGTCAACGAGATTTTGAATATATGATCCATC 458
Db      93 GlnGlnLysValLeuSerGlnPheGlnAla---LysAspValCysGlnAlaTrpProSer 111
QY      459 CCTCTGATGTGGAGAGAGATTAAGCTTGAATGGCCCTACAGAGCAATTGTACTTC 518
Db      112 PheAlaValAlaProLysLeuLeuThrGlnTrpProLeuGlnGlnLysLeuLeu 131
QY      519 CGAAGACACTACCAAGAAAGTTCAAGAACAGACCTTCAAGAAAGCTCTGTAATGTG 578
Db      132 AlaGlnHisSerTrpLysHisLysAsnThrTrpIleThrArgSerThrLeuAspIle 151
QY      579 CTATACACTTTTGGCAAGCGCTCAGACCTGACTGATCTTTGGCTTAATCGTTATTA 638
Db      152 LeuHisThrPheAlaSerSerSerGlyPheArgLysValPheGlyLeuAsnAlaLeuLeu 171
QY      639 AGAAGACAGATTTTGCAGTGGCAAGCTTCTAATGCTCAAGTCTCTCGACTACTGCTCT 698
Db      172 ArgArgGlnGlyLeuGlnTrpAspSerSerAsnAlaLysGlnLeuGlyTyrCysAla 191
QY      699 TCAGAGGGGTATTAACATTTCTTGGGAATAGGCAATGAACCTTAACATTTCTTGAAG 758
Db      192 GlnArgSerTyrAsnIleSerTrpGlnLeuGlnProAsnSerPheArgLysLys 211
QY      759 GCTGATTTTTCATCAATGGGTCGAGCTTGAAGAAAGTATTAATCAATTCATTAACCT 818
Db      212 SerGlyIleCysIleAspGlyPheGlnLeuGlyArgAspPheValHisLeuArgGlnLeu 231
QY      819 CTA---AGAAAGTCCACCTTCAAAATGCAAAACTCTAATGCTCGATGTTGGTCAAGCT 875
Db      232 LeuSerGlnHisProLeuTyrArgHisAlaGlnLysLeuTyrGlyLeuAspValGlyGlnPro 251
QY      876 CGAAGAAAGACGGCTAAGATCTGAAGAGCTTCTGAAGGCTGTGGAGAAAGTATGAT 935
Db      252 ArgLysHisThrGlnHisLeuLeuArgSerPheMetLysSerGlyGlyLeuAlaIleAsp 271
QY      936 TCAAGTTCATGCGACATCACTATTTGAAGAGCAAGCTGCAACGAGAAAGATTTTCTA 995
Db      272 SerValThrTrpHisHisTyrTyrValAsnGlyArgSerAlaThrAspGlnAspPheLeu 291
QY      996 AACCTGATGATTTGACATTTTATTCATCTGTGCAAAAGTTTTCAGGTGGTGTAG 1055
Db      292 SerProGlnValLeuAspSerPheAlaThrAlaIleHisAspValLeuGlyIleValGln 311
QY      1056 AGCACCAAGCGCTGCAAGAGGCTGTGTTAGAGAAACAAGCTTGCATATGAGGCGCA 1115
Db      312 AlaThrValProGlyLysLysValTrpLeuGlnGlnThrGlySerAlaTyrGlyGly 331

```

```

QY 1116 GCCCCTGCTGATCCGACACCTTGGACCTGGCTTTATGTGGCTGATTAATAATGGGCTTG 1175
DB 332 AAlaProGlnLeuSerAenThrTyValAlaGlyPheMetCrrLeuAapLysLeuGlyLeu 351
QY 1176 TCAGCCCGAATGGGAATAGAGAGTGTGATGAGCAATATCTTTGGAGAGAAATAC 1235
DB 352 AAlaAlaArgArgGlyGlyLeuAspValValMetCArgGlnAlaSerPheGlyAlaGlySerTyr 371
QY 1236 CATTAGTGGATGAGAAACCTTCGATCCTTTTACCTGATATTGGCTATCTCTTGGTTCAG 1295
DB 372 HisLeuValLeuPheAlaGlyPheLysProLeuProAspTyrTrpLeuSerLeuLeuTyrLys 391
QY 1296 AAATGTGGGACCAAGGTGTTATATGGCAAGCTGCAAGTTCAAGAGAGAGAGCTT 1355
DB 392 ArgLeuValGlyTrpArgValLeuGlnAlaSerValGlnAlaAspAlaArgArgPro 411
QY 1356 CGAGTAACTCTTACCTTGACAAAACACGACATCCAGATTAAGAGAGAGATTTAACT 1415
DB 412 ArgValTyrLeuHisCysThrAsnProArgHisProLysTyrArgGlnGlyAspValThr 431
QY 1416 CTGTATGCTAAACCTCCATACGTCACCAAGTACTTGGGTTACCTTATCCTTTTCT 1475
DB 432 LeuPheAlaLeuMetLeuSerAenValThrGlnSerLeuGlnLeuProLysGlnLeuTyr 451
QY 1476 AACAGCAAGTGGATAAATACCTTCTTAAGACCTTGGGACCTCAGATTACTTCCAA 1535
DB 452 SerLysSerValAspGlnTyrLeuLeuLeuProHisGlyAspSerIleLeuSerArg 471
QY 1536 TCTGTCAACTCATGTGCTTAACCTTAAGATGGTGGATGATCAACCTTGGCCACCTTA 1595
DB 472 GluValGlnLeuMetGlyAspGlyLeuGlnMetValAspAspGlnTrpLeuProAlaLeu 491
QY 1596 ATGGAATAAACCTCTCCGCGACGAGAGTTCACCTGGGCTTGCAGCTTCTCATATAGTTT 1655
DB 492 HisGlnMetAlaLeuAlaProGlySerThrLeuGlyLeuProAlaPheSerTyrGlyPhe 511
QY 1656 TTTGTGATAAGAAATGCCAAAGTTGCTGCTTGCATC 1691
DB 512 TyrValIleArgAsnAlaLysAlaIleAlaCyste 523

```

RESULT 7

```

Q9HB37 PRELIMINARY; PRT; 592 AA.
AC Q9HB37;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Heparanase-like protein HPA2c.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20483645; PubMed=11027606;
RA McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RA Hirschcock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.;
RT "Cloning and Expression Profiling of Hpa2, a Novel Mammalian
RT Heparanase Family Member.";
RL Biochem. Biophys. Res. Commun. 276:1170-1177(2000).
DR EMBL:AF282887; AAC23423.1;
DR GO; GO:0005622; C:intracellular; TAS.
DR GO; GO:0030305; F:heparanase activity; TAS.
DR InterPro; IPR005199; Glyco_hydro_79N.
DR Pfam; PF03662; Glyco_hydro_79n; 1.
SQ SEQUENCE 592 AA; 66580 MW; 95C384AD9A74258E CRC64;

```

Alignment Scores:

```

Pred. No.: 4.67e-85 Length: 592
Score: 1154.50 Matches: 250
Percent Similarity: 57.84% Conservative: 82
Best Local Similarity: 43.55% Mismatches: 189
Query Match: 37.02% Indels: 53

```

```

DB: 4 Gaps: 9
US-10-676-079-3 (1-1721) x Q9HB37 (1-592)
QY 120 CCGCTGGTCCCTCCCTCCCTGGCGCCCTG----- 149
DB 18 ProProAlaCysLeuAlaProGlyAlaLeuTyrLeuAlaLeuLeuLeuHisLeuSerLeu 37
QY 150 -----CCCGACCTGCG-----CAAGCA 167
DB 38 SerSerGlnAlaGlyAspArgArgProLeuProValAspArgAlaAlaGlyLeuLysGlu 57
QY 168 CAGAGCTGTGGACCTGACCTTCTTACCCAGAGACCCGCTGCACCTGGTACCCCTG 227
DB 58 LysThrLeuIleLeuLeuAspValSerThrLysAsnProValArgThrValGlnGluAsn 77
QY 228 TTTCTTCGCTCCATCCATTAAGCCCAACCTGGCCAGGACCCGGGCTTCATCCCTCG 287
DB 78 PheLeuSerLeuGlnLeuAspProSerIleIleHisAsp--GlyTrpLeuAspPheLeu 96
QY 288 GGTTCCTCAAAAGCTTGTGTAACCTTGGCCAGAGGCTTGTCTGCTGACCTGAGTTGGT 347
DB 97 SerSerLysArgLeuValThrLeuAlaArgGlyLeuSerProAlaPheLeuArgPheGly 116
QY 348 GGCACCAAGACAGACTTCTTAATTTTC-----GATCCCAAGAGATCAACC 395
DB 117 GlyLysArgThrAspPheLeuGlnPheGlnAsnLeuArgAsnProAlaLysSerArgGly 136
QY 396 TTGAAGAGAGAAAGTAACTAGGCAATCTCAAGTCAACAGGATTA----- 440
DB 137 GlyProGlyProAspTyrTyrLeuLysAsnTyrGluAspAspIleValArgSerAspVal 156
QY 441 -----TGCAATATGATTCATCCCTCTGATGTGAGAGAG 479
DB 157 AlaLeuAspLysGlnLysGlyCysLysIleAlaGln--HisProAspValMetLeuGlu 175
QY 480 TTACGGTGGAAATGGCCCTTACAGAGCAATTG--CTACTCGAGAAACATACCGAAA 536
DB 176 LeuGlnArgGlyLysAlaAlaGlnMetCHisLeuValLeuLeuLysGlnIlePheSerAsn 195
QY 537 AAGTCAAGAACGACCTACTCAAGAGCTGTGATGTGCTATACCTTTTGGCAAC 596
DB 196 ThrTyrSerAsnLeuIleLeuThrAlaArgSerLeuAspLysLeuTrpAsnPheAlaAsp 215
QY 597 TGTCTGAGACTGGACCTTGAATCTTTGGCTTAATGCTTATTAAGAACAGACAGATTGGAC 656
DB 216 CysSerGlyLeuHisLeuIlePheAlaLeuAsnAlaLeuArgAsnProAsnAsnSer 235
QY 657 TGGACAGTTCTAATGCTCAGTTGCTCTGAGCTACTGCTTCCAAAGGCTTAACAT 716
DB 236 TrpAsnSerSerSerAlaLeuSerLeuLeuLysTyrSerAlaSerLysLysTyrAsnIle 255
QY 717 TCTTGGGAATAGGCAATGAACTTAACAGTTTCCTTAAGAGGCTGATTTTCATCAAT 776
DB 256 SerTrpGlnLeuGlnLysAsnGlnProAsnAsnTyrArgThrMetCHisGlyArgAlaValAsn 275
QY 777 GGGTTCAGTTAGAGAGATTAATTAATTCATTCATTAACCTTCAAGAAAG--TCACAC 833
DB 276 GlySerGlnLeuGlyLysAspTyrIleGlnLeuLysSerLeuLeuGlnProIleArgIle 295
QY 834 TTCAAAAATGCAAAACCTATGCTCTGATGTTGTGCTGACGCTTGAAGAAAGAGCGCTAAG 893
DB 296 TyrSerArgAlaSerLeuTyrGlyProAsnIleGlyArgProArgLysAsnValIleAla 315
QY 894 ATGCTGAAGAGCTTCTGAAAGGCTGGTGAAGAGATGATGATTCATGATGATGATC 953
DB 316 LeuLeuAspGlyPheMetCysValAlaGlySerThrValAspAlaValThrTrpGlnHis 335
QY 954 TACTATTGATAGAGAGAGCTGCTACAGAGAGATTTTCTAAACCTGATGATTTGAC 1013
DB 336 CysTyrIleAspGlyArgValValLysValMetAspPheLeuLysThrArgLeuLeuAsp 355
QY 1014 ATTTTATTTCATCTGTGCAAAAGTTTTCACAGTGTGTTGAGAGCACAGGCTGGCAAG 1073

```

```

Db      356 ThrLeuSerAspGlnIleArgLysIleGlnLysValIleAsnThrTyrThrProGlyLys 375
      1074 AAGGCTGGTGGAGAAACAAGCTGTGATATGAGAGCGCCCTTGCTATCCGAC 1133
      376 LysIleTyrLeuGlnGlyValValThrThrSerAlaGlyGlyThrAsnAsnLeuSerAsp 395
      1134 ACCCTTGCAGCTGGCTTTATGTGCTGATTAATGGCGCTGACGCCGAATGGAGATA 1193
      396 SerTyrAlaIleGlyPheLeuThrLeuAsnThrLeuGlyMetLeuAlaAsnGlnGlyIle 415
      1194 GAAGTGTGATGAGCAAGCAATATTCTTTGAGCAGCAAACTACCATTTAGTGAATAAAC 1253
      416 AspValIleAlaArgHisSerPheAspAsnIleGlyTyrAsnHisIleValAspGlnAsn 435
      1254 TTGCAATCCTTACCTGATTAATGCTATGCTTCTGCTTCAAGAAATGGTGGCAGCAAG 1313
      436 PheAsnProLeuProAspTyrThrLeuSerLeuLeuTyrLysArgLeuIleGlyProLys 455
      1314 GTGTTAATGGCAAGCGTGCAGAGTTCAAGAGAAAG----- 1349
      456 ValLeuAlaValHisValIleGlyLeuGlnArgLysProArgProGlyArgValIleArg 475
      1350 ---AAGCTTCCAGATATCTTCAATGTCACAAACATGCAATCCAAAGTAAAGAGGA 1406
      476 AspLysLeuArgIleTyrAlaHisCysThrAsnHisIleAsnHisValAsnTyrValArgGly 495
      1407 GATTTAATCCTGTGACCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1466
      496 SerIleThrLeuPheIleIleAsnLeuHisArgSerArgLysLysIleLysLeuAlaGly 515
      1467 CTTTTTTTCAACAGCAAGTGAATAATCACTTCTTAAGACCTTTGGACCTGATGATTA 1526
      516 ThrLeuArgAspLysLeuValHisGlnTyrLeuLeuGlnProTyrGlyGlnGlyLeu 535
      1527 CTTTCCAAATCTGTCCAACTCAATGCTTAATCTTAAGATGCTGATCAAACTTTG 1586
      536 LysSerLysSerValGlnLeuAsnGlnGlnProLeuValMetValAspAspGlyThrLeu 555
      1587 CCACCTTTAATGGAATAAACTCTCCGCCAGAGAGTTCCTGGGCTTGGCAGCTTTCTCA 1646
      556 ProGlnLeuLysProArgProLeuArgAlaGlyArgThrLeuValIleProValThr 575
      1647 TATAGTTTTTTGTGATAGAATGCCAAAGTTGCTGCTGC 1688
      576 MetGlyPhePheValValLysAsnValAlaLeuAlaCys 589

RESULT 8
Q8WMQ2 PRELIMINARY; PRT; 592 AA.
AC Q8WMQ2;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DB Heparanase 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Peesque Safontas B.J.O.P.S.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ299719; CAC82491.1; -
DR Genew; HGNC:18374; HPS2.
DR Interpro; IPR005199; Glyco_hydro_79N.
DR Pfam; PF03662; Glyco_hydro_79n; 1.

```

```

SQ SEQUENCE 592 AA; 66520 MW; 9478841FEACD558B CRC64;
Alignment Scores:
Pred. No.: 2,09e-84 Length: 592
Score: 1146.50 Matches: 249
Percent Similarity: 57.67% Conservative: 82
Best Local Similarity: 43.38% Mismatches: 190
Query Match: 36.76% Indels: 53
DB: 4 Gaps: 9

US-10-676-079-3 (1-1721) x Q8WMQ2 (1-592)
QY CCGCTGGGCTCCCTCTCCCTGGCGCCCTG----- 149
DB 18 ProProAlaCysLeuAlaProGlyAlaLeuTyrLeuAlaLeuLeuHisIleuSerLeu 37
QY 150 -----CCCGACCTGCG-----CAAGCA 167
DB 38 SerSerGlnAlaGlyAspArgArgProLeuProValAspArgAlaAlaGlyLeuLysGlu 57
QY 168 CAGGACGTGCGGAGCTGAGCTTCTTCAACCGAGCGCGTGCACCTGTGAGCCCTCG 227
DB 58 ThrThrLeuIleLeuLeuAspValSerThrLysAsnProValArgThrValAsnGlyAsn 77
QY 228 TTCTGTGCTGCACCATTCAGCCCAACCTGGCCAGCAGACCGCGGTTCTCATCTCTG 287
DB 78 PheLeuSerLeuGlnLeuAspProSerIleIleHisAsp---GlyTyrLeuAspPheLeu 96
QY 288 GGTTCCTCAAACTTCGATCTTGGCCAGAGGCTTGTCTCTGCTGATCACTGAGTTGGT 347
DB 97 SerSerLysArgLeuValThrLeuAlaArgGlyLeuSerProAlaPheLeuArgPheGly 116
QY 348 GGCACCAAGACAGACTTCTCAATTTT-----GATCCCAAGAGAAATCAACC 395
DB 117 GlyLysArgThrAspPheLeuGlnPheGlnAsnLeuAlaGlnAsnProAlaLysSerArgGly 136
QY 396 TTTGAAGAGAGAACTTACTGGCAATCTCAAGTCAACAGATATT----- 440
DB 137 GlyProGlyProAspTyrTyrLeuLysAsnTyrGlnAspAspIleValArgSerAspVal 156
QY 441 -----TGCAAATATGAGATCCATTCCTCTGATGTGAGAGAG 479
DB 157 AlaLeuAspLysGlnLysGlyCysLysIleAlaGln---HisProAspValMetLeuGlu 175
QY 480 TTACGCTTGGATGCGCCCTTACCAAGAGCAATTG---CTACCTCCAGAACACTACCAAGAA 536
DB 176 LeuGlnArgGlnLysAlaAlaGlnMetCHisLeuValIleLeuLysGlnGlnPheSerAsn 195
QY 537 AAGTTCAGAACAGCACTTACTCAAGAAAGCTGTGAGATGTGCTATACACTTTTCAAC 596
DB 196 ThrTyrSerAsnLeuIleLeuThrAlaArgSerLeuAspLysLeuTyrAsnSerAlaAsp 215
QY 597 TGTCTCAGAGACTGAGACTTGTGCTTTGGCTTAATGCGTTATTAAGAACAGCAAGTTGCG 656
DB 216 CysSerGlyLeuHisIleLeuIlePheAlaLeuAsnAlaLeuArgAsnProAsnAsnSer 235
QY 657 TGGACACTTCTTAAGTCAAGTGTCTCTGAGACTATCTCTGCAAGGAGTAATCAATT 716
DB 236 TrpAsnSerSerSerAlaLeuSerLeuLeuLysTyrSerAlaSerLysLysTyrAsnIle 255
QY 717 TCTTGGAAGTACGAGCAATGAACCTTAACAGTTTCTTAAAGAGCGTGAATTTTCATCAT 776
DB 256 SerTyrGlnLeuGlnLysAsnGlnProAsnAsnTyrArgThrMetHisGlyArgAlaValAsn 275
QY 777 GGGTTCGAGTTGAGAGAAATTAATTCAATTGCAATAACTTTAAGAAAG---TCCACC 833
DB 276 GlySerGlnLeuGlyLysAspTyrIleGlnLeuLysSerLeuLeuGlnProIleArgIle 295
QY 834 TTCAAAATATGCAAACTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 893
DB 296 TyrSerArgAlaSerLeuTyrGlyProAsnIleGlyArgProArgLysAsnValIleAla 315
QY 894 ATGCTGAAGAGCTTCTGAAAGCTGTGAGAGAAAGTATGATTCAGTTACATGAGCATCAC 953

```

```

Db      316 LeuLeuaspGlyPheMetLysValAlaGlySerThrValAspAlaValThrProGlnHis 335
QY      954 TACTATTGATGAGACCGAGCTGTACACAGGAGATTCTTAAACCTGTATGTTAGAC 1013
Db      336 CyetYrLleAerGlyArgValAlaValMetAspPheLeuYsrThrArgLeuLeuAsp 355
QY      1014 ATTTTATTTTCATCTGTGCAAAAAGTTTCCAGGTGTGAGACACCGAGCTGGCAG 1073
Db      356 ThrLeuSerAspGlnIleArgLysIleGlnValAlaValAsnThrYrThrProGlyLys 375
QY      1074 AGGTCTGTAGGAGAAACAGCTGTGATGAGAGCGAGCGCTGTATCGGAC 1133
Db      376 LysIleThrPheGlnGlyValAlaThrThrSerAlaGlyGlnThrAsnLeuSerAsp 395
QY      1134 ACCTTTCAGACTGGCTTTATGTGGCTGATTAATTTGGGCTGTGACCGCAATGGAGATA 1193
Db      396 SerYrAlaAlaGlyPheLeuThrPheLeuAsnThrLeuGlyMetLeuAlaAsnGlnGlyIle 415
QY      1194 GAAGTGTGATGAGGCAAGTATTTCTTGGACAGGAACTACCTTTAGTGTGAGAAAC 1253
Db      416 AspValValIleArgHisSerPheAspPheIleGlyTrpAsnHisLeuValAspGlnAsn 435
QY      1254 TTGGATCTTATCCTGATTTATGCTATCTCTGTTCAAGAAATTTGGGGGAGCCAG 1313
Db      436 PheAsnProLeuProAspYrThrPheSerLeuLeuYrLysArgLeuIleGlyProLys 455
QY      1314 GTGTAAATGCGAACGCTGCAAGGTTTCAAGAGAGG----- 1349
Db      456 ValLeuAlaValHisValAlaGlyLeuGlnArgYsrProArgProGlyArgValIleArg 475
QY      1350 ---AGCTTGTAGATATCTCTCATTTGCAAAACATGACATCCAGATTAAGAGGA 1406
Db      476 AspLysLeuArgIleYrAlaHisCysThrAsnHisIleAsnHisAsnYrValArgGly 495
QY      1407 GATTTAACCTGTGATGCGATTAACCTGTACAGTCAACAGTACTGGCGGTATACCTAT 1466
Db      496 SerIleThrLeuPheIleIleAsnLeuHisArgSerArgLysValIleLysLeuAlaGly 515
QY      1467 CTTTTTCTAACAACAGATGATTAACCTTTCTAAGACCTTGGAGCTCATGATTA 1526
Db      516 ThrLeuArgAspLysLeuValHisGlnIleLeuLeuGlnProYrGlyGlnGlnGlyLeu 535
QY      1527 CTTTCCAAATCTGTCCCACTCAATGTGTCTAATCTTAAGATGTGGATGATCAAACTTG 1586
Db      536 LysSerLysSerValGlnLeuAsnGlyGlnProLeuValMetValAspAspGlyThrLeu 555
QY      1587 CCACTTTAATGGAACAACTCTCCGCGCAAGAAAGTTCAGGCGCTTGCAGCTTCTCA 1646
Db      556 ProGlnLeuYsrProArgProLeuArgAlaGlyArgThrLeuValIleProValThr 575
QY      1647 TATAGTTTTTTGTGATGAGAAATGCAAGTTGCTTGC 1688
Db      576 MetGlyPhePheValValLysAsnValAlaAsnAlaLeuLeuLys 589

RESULT 9
Q8WMQ1 PRELIMINARY; PRT; 548 AA.
AC 08WMQ1;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-UTN-2003 (TEMBLrel. 24, Last annotation update)
DE Heparanase 3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Peasegue Safoncas B.J.O.P.S.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Legoux P., Legoux R., O'Brien D., Salome M.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; A0299720; CAC82492.1; -
DR InterPro; IPR005139; Glyco_hydro_79N.
DR Pfam; PF03662; Glyco_hydro_79n; I.
SQ SEQUENCE 548 AA; 61771 MW; B8986303FC73A60A CRC64;

Alignment Scores:
Pred. No.: 9, 08e-74 Length: 548
Score: 1015.50 Matches: 224
Percent Similarity: 56.53% Conservative: 79
Best Local Similarity: 41.79% Mismatches: 180
Query Match: 32.56% Indels: 53
DB: Gaps: 9

US-10-676-079-3 (1-1721) x Q8WMQ1 (1-548)
QY 84 GCGCTGCCCGCGCGCTGATGCTGCTGCGGCGCGCTGCGCTTCCCTCCGCGC 143
Db 6 AlaPheProGlnAlaMetLeuSerSerAsnSerArgProProAlaCysValAlaProGly 25
QY 144 GCGCTG----- 149
Db 26 AlaLeuYrLeuAlaLeuLeuLeuHisLeuSerLeuSerSerGlnAlaGlyAspArg 45
QY 150 CCGCGACTGCG-----CAAGCACAGAGAGCTGCGACCTGACCTTTC 191
Db 46 ProLeuProValAspArgAlaAlaGlyLeuYrGlnYrThrLeuIleLeuLeuAspVal 65
QY 192 TTCACCCAGAGGCGCTGCACTGCTGAGCCCTGCTTCTGCTGCTGACCATGAGCC 251
Db 66 SerThrLysAsnProValArgThrValAsnGlnAsnPheLeuSerLeuGlnLeuAspPro 85
QY 252 AACCTGCGCACAGGACCGCGGCTTCATCTCTGCTGCTTCCAAAGCTTGACCTTG 311
Db 86 SerIleIleHisAsp---GlyTrpLeuAspPheLeuSerSerLysArgLeuValThrLeu 104
QY 312 GCCAGAGGCTTGTCTCTGCTGATGCTGAGCTTGTGTGACCAAGACAGACTTCCTAAT 371
Db 105 AlaArgGlyLeuSerProAlaPheLeuArgPheGlyGlyLysArgThrAspPheLeuGln 124
QY 372 TTC-----GATCCCAAGAAAGAAATCACTTGAAGAGAAATTAATGCGCAA 419
Db 125 PheGlnAsnLeuArgAsnProAlaLysSerArgGlyGlyProGlyProAspYrYrLeu 144
QY 420 TCTCAAGTCAACCGAGATTT-----TGC 443
Db 145 LysAsnYrGlnAspAspIleValAlaArgSerAspValAlaLeuAspLysGlnLysGlyCys 164
QY 444 AATATGATTCATCCCTCTGATGTGAGAGAAAGTACGTTGGAATGAGCCCTACAG 503
Db 165 LysIleAlaGln---HisProAspValMetLeuGlnLeuGlnArgLysValAlaGln 183
QY 504 GAGCAATTG---CTACTCCGAGAAACATCAAGAAAGATTCAAGAACAGCACTTACTA 560
Db 184 MetHisLeuValLeuLeuLysGlnGlnPheSerAsnThrYrSerAsnLeuLeuLeu 203
QY 561 AGAAGCTGTGATGATGCTATACATTTTGCAAAATGCTGCGAGACTGACCTGATCTT 620
Db 204 AlaArgSerLeuAspLysLysLysYrAsnSerAlaAspCysSerGlyLeuHisLeuIlePhe 223
QY 621 GGCCTAATGCGTTATTATTAAGAACAGACAGATTTGACGTGAGACAGTTTATGCTCAGTT 680
Db 224 AlaLeuAsnAlaLeuArgAsnProAsnAsnSerTrpAsnSerSerSerAlaLeuSer 243
QY 681 CTCTGAGTACTGCTCTTCCAGAGGGGATTAACATTTCTTGGAACTAGCAATGAACT 740
Db 244 LeuLeuLysYrSerLysSerLysYrLysValIleSerTrpGlnLeuGlnGlnGlyPro 263
QY 741 AACAGTTTCTTAAAGAGCTGATATTTTCATCAATGCGTCCAGTTAGAGAAAGTTTAT 800

```


Db	238	TyrSerArgAlaSerLeuIyrG1gIyProAlnIleG1yAArgProArgLysValIleAla	257
Qy	894	ATGCTGAGAGAGCTTCTGTAAGGCTGTGTGGAGAAGTGAATTCATTCAGTAACGAGATAC	953
Db	258	LeuLeuAerGlyPheMetLysValAlaGlySerThrValAspAlaValThrTrpGlnHis	277
Qy	954	TACTATTTGGAATGGAACGAGACTGCTCAAGGAGAAATTTTCTAAACCTGATATGATGGAC	1013
Db	278	CysTyrXileAspGlyArgValValLysValMetAspPheLeuIyrThrArgLeuLeuSer	297
Qy	1014	ATTTTATTTATTCATCTGTGCAAAAGATTTTCCAGGTGGTTGAGAGACCAAGGCTGGCAAG	1073
Db	298	ThrLeuSerAspGlnIleArgLysIleGlnIyrValIleValAsnThrTyrThrProGlyLys	317
Qy	1074	AAGGTGTGGTTGAGAGAAACAAGCTGTGCATGTGAGGCGGAGCCCTTGCTATCCGAC	1133
Db	318	LysIleTrpLeuGlnGlyValValIleThrTrpSerAlaGlyGlyThrAsnAsnLeuSerAsp	337
Qy	1134	ACCTTTGACGCTGCTGTTATATGTGGCTGGAATAATTGGGCTGTGACCCGAAATGGAAAT	1193
Db	338	SerTyrAlaAlaGlyPheLeuTrpLeuAsnThrLeuGlyMetLeuAlaAsnGlnIyrIle	357
Qy	1194	GAGGTGGTATGAGGCAAGATATCTTTTGGAGAGAAATCCATTTATGTGGATGAAGAC	1253
Db	358	AspValValIleAsnHisSerPhePheAsnHisGlyTyrAsnHisLeuValAspGlnIleAsn	377
Qy	1254	TTGCATCCTTAACTGATTAATTTAGGCTATCTTCTGTTCAGAAATTTGGTGGGACCAAG	1313
Db	378	PheAsnProLeuProAspTyrTrpLeuSerIleLeuTyrLysArgLeuIleGlyProLys	397
Qy	1314	GTGCTTAATGGCAAGGCTGCAAGCTTCAAGAAGAGG-----	1349
Db	398	ValLeuAlaValAlaHisValAlaGlyLeuGlnAlaGlyLysProArgProGlyArgValIleArg	417
Qy	1350	---AAGCTTGCAGTATACCTTCAATTCAGCAAAACAGTCAATCCAAAGGTATTAAGAAGA	1406
Db	418	AspLysLeuAspArgLysIleTyrAlaHisSerGlyThrAsnHisValAsnHisValAsnTyrValArgGly	437
Qy	1407	GATTTAACTCTGTATGCTCCATAAATCTCCATAACGTGCACCAAGTACTTGGCTGCTTAT	1466
Db	438	SerIleThrTrpPheIleIleAsnLeuHisArgSerArgLysLysIleLeuLeuAlaGly	457
Qy	1467	CCTTTTCTTAACAACAGATGGATTAATATCCTTCTTAAGACCTTTGGGACCTCATGAGATTA	1526
Db	458	ThrLeuAspAspLysLeuValHisSerGlnTyrLeuLeuGlnProTyrGlyGlnGlyLeu	477
Qy	1527	CTTTGCAATCTGTGCAATCTCAATGCTGTAACTTTAAAGATGTGGATGATCAACCTTGG	1586
Db	478	LysSerLysSerValGlnLeuAsnGlyGlnProLeuValMetValAspAspGlyThrLeu	497
Qy	1587	CCACCTTTAATGGAAAAACCTTCGGGCGACGAAGTCACTGAGGCTTGGCCAGCTTTCTCA	1646
Db	498	ProGlnLeuLysProArgProLeuAspArgAlaGlyArgTrpLeuValIleLeuProValThr	517
Qy	1647	TATAGTCTTTTGTGATAAGAAATGCCAAAGTGTGCTGCTGC	1688
Db	518	MetGlyPhePheValValLysAsnValAsnAlaLeuAlaCys	531
RESULT 11			
Q9HB39 PRELIMINARY; PRT; 480 AA.			
AC	Q9HB39	01-MAR-2001 (Tremblrel. 16, Created)	
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)		
DT	01-JUN-2003 (Tremblrel. 24, Last annotation update)		
DE	Hepatitis-like protein HPA2a.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
NP	[1]		
SEQUENCE FROM N.A.			

RX	MEDLINE=20483645; PubMed=11027606;
RA	McKenzie E., Tyson K., Stamps A., Smith P., Turner P., Barry R.,
RB	Hancock M., Patel S., Barry E., Stubberfield C., Terrett J., Page M.,
RT	"Cloning and Expression Profiling of Hpa2, a Novel Mammalian
RT	Hepatocarcinoma Family Member.";
RL	Biochem. Biophys. Res. Commun. 276:1170-1177 (2000).
DR	PIR: JCF506; JCF506.
DR	InterPro: IPR005199; Glyco hydro 79N.
DR	Pfam: PF03662; Glyco hydro 79n; I-
SQ	SEQUENCE 480 AA; 5390 Km; F75F89SF67ACLF83 CRC64;
Alignment Scores:	
Pred. No.:	3..4e-64 Length: 480
Score:	897.50 Matches: 202
Percent Similarity:	49.20% Conservative: 74
Best local Similarity:	36.01% Mismatches: 146
Query Match:	28.78% Indels: 139
Dbl:	4 Gaps: 9
US-10-676-079-3 (1-1721) x Q9HBJ9 (1-480)	
OY	120 CGCGTGGGTCCTCCCTGCCGTGGGCCTG-----149
DB	18 ProProAlaCysLeuAlaProGlyAlaLeuTyrlleuAlaLeuLeuHisIleuSerLeu 37
OY	150 -----CCCGACCTGGC-----CAACA 167
DB	38 SerSerGIlaAlaGlyAbaPrArgArgProIleuProValaAbaArgAlaAlaGlyLeuLyGlu 57
OY	168 CAAGAGCGTGCGAGCCTGGACTTTTCAACCAGAGCCGCGTAGCACTGGAGCCCTCG 227
DB	58 LysThrLeuIleuLeuAspAlaSerThrLysaenProValaArgThyValaAngLaaen 77
OY	228 TTCTGTCCGTACCATTCAGCCCAACCTGGCCAGCAGCACCGCGGTTCTCATCTCTCG 287
DB	78 PhenSerIeuGlneuhSpserIerleiIehIsaP---GLYTrpleuaPhneu 96
OY	288 GGTTTCCAAAGCTTGATCTTGGCCAGAGCTTGTCTCTGGCTAACCTGAGTTGGT 347
DB	97 SerSerIyaIrghIeuValaThrIeuAlaArgGlyIeuSerProAlaPhelaeArgPhGly 116
OY	348 GGCACCAAGCAGACTTCTTAATTTTC-----GATCCCAAGAAGAATCAACC 395
DB	117 GLYLysaIrghThraPhneuGlInPhelGlInbenLeuArghaenProAlaLysSerArg-- 135
OY	396 TTTGAAGAGAAATTACTGCAATCTCAAGTCAACCAAGATATTGCAATAATGATCC 455
DB	136 -----GLYgly 137
OY	456 ATCCCTCCGATGTGGAGAGAAAGTTAAGCTTGAATGGCCCTTACAGAGCAATTGCTA 515
DB	138 ProGlyProasp-----141
OY	516 CTCGAGAAACACTACCAAAAAGTTTCAAGAAGCAACTTACAAAGAGCTGTAGAT 575
DB	142 -----TyrrTyrlleuYhaanTyrlu----148
OY	576 GTGCTATACCTTTTGCAAATGCTCAGAGACTGACCTTGATCTTTGGCTTAATGCGTTA 635
DB	148 -----148
OY	636 TTTAAGAACAGAGATTGACAGTGAACAGTTCTAATGCTCAAGTGTCTCGAATGCG 695
DB	148 -----148
OY	696 TCTTCCAAAGGGTATTAACATTTCTTGGAACTAGGCAATGAACCTTAAGTTCTTAAg 755
DB	149 -----AepgluProdaanaenTyArgThr 156
OY	756 AAGCGTGAATTTTTCATCAATGGGTCGAGTTAGAGAAAGTTTATTCATTAATGCATAAA 815
DB	157 MechIeglYArgAlaValaIsmelYserIeuIeuGlyYsaByrYtlIegIlnIeuYser 176

QY 816 CTTCTAAGAAAG---TCCACCTTCAAAAAATGCAAACTATAGTCTGATGGTGC 872
 DB 177 LeuLeuGlnProIleArgIleTyrSerKhaIasleuTyrGlyProAsnIleGlyArg 196
 QY 873 CCTCGAAGAAAGACGGCTTAAGATGCTGAAGACCTTCTGAAAGCGTGGAGAACTGAT 932
 DB 197 ProArgLysAsnValIleAlaLeuLeuAspGlyPheMetLysValAlaGlySerThrVal 216
 QY 933 GATTGAGTTACATGCGATCCTACTATTTGATGACCGACCTGCTACCAAGGAATTTT 992
 DB 217 AspaIaValThrTrpGlnHisCysGlyTyrIleAspGlyArgValValLysValMetAspPhe 236
 QY 993 CTAAACCTGATGTGATGATGATTTTATTTATTCATCTGCAAAAAGTTTTCAGAGTGT 1052
 DB 237 LeuLysThrArgLeuLeuAspThrLeuSerAspGlnIleArgLysIleGlnLysValVal 256
 QY 1053 GAGACACACCGCTGGCAAGAGCTGTGTTAGAGAAACAGCTCTGCATATGAGAGC 1112
 DB 257 AsnThrTyrThrProGlyLysLysIleTrpLeuGlnGlyValValThrThrSerAlaGly 276
 QY 1113 GAGAGCGCCCTGCTATCCGACACCTTGGACCTGCTTATATGCTGATTAATTTGGGC 1172
 DB 277 GlyThrAsnAsnLeuSerAspSerTyrAlaAlaGlyPheLeuTrpLeuAsnThrLeuGly 296
 QY 1173 CTGTGAGCCCGAATGGAATAGAGTGTGATGAGGCAAGTATTTCTTGGACAGGAAC 1232
 DB 297 MetLeuAlaAsnGlnGlyIleAspValValIleArgHisSerPhePheAspHisGlyTyr 316
 QY 1233 TACCATTTAGTGAATGAAATCTGATCCTTACCTGATTAATGCTATCTTCTGTTTC 1292
 DB 317 AsnHisLeuValAspGlnAsnPheAsnProLeuProAspTyrTrpLeuSerLeuLeuTyr 336
 QY 1293 AAGAAATTTGGTGGGACCAAGGTGTTAATGCAAGCTGCAAGCTTCAAGGAAG--- 1349
 DB 337 LysArgLeuIleGlyProLysValIleuAlaValHisValAlaGlyLeuGlnArgLysPro 356
 QY 1350 -----AAGCTTGGATGATCTTCACTTCACTTGGCAAAACTGAC 1385
 DB 357 ArgProGlyArgValIleArgAspLysLeuArgIleTyrAlaHisCysThrAsnHis 376
 QY 1386 AATCCAAAGTATTAAGAGAGATTTAACTGTGATGCAATAAAGCTCCATTAAGCTCAC 1445
 DB 377 AsnHisAsnTyrValArgGlySerIleThrLeuPheIleIleAsnLeuHisArgSerArg 396
 QY 1446 AAGTACTCGGCTACCTTATCTTTTCTTAACAGCAAGTGGATTAATACCTCTTAAGA 1505
 DB 397 LysLysIleLysLeuAlaGlyThrLeuArgAspLysLeuValHisGlnTyrLeuLeuGln 416
 QY 1506 CCTTTGGGACCTCAGATTAATCTTCAATCTGTCACATCAATGGTCTAACTGAAG 1565
 DB 417 ProTyrGlyGlnGlyLeuLysSerLysValGlnLeuAsnGlyLysProLeuVal 436
 QY 1566 ATGTGATGATCAAACTTGGCACTTAAATGAGAAACCTCTCCGCGCAGAGAACTTCA 1625
 DB 437 MetValAspArgGlyThrLeuProGlyLeuLysProArgProLeuArgLysAlaGlyArgThr 456
 QY 1626 CTGGCGCTTCCAGCTTCTCATATAGTTTGTGATTAAGAAATGCCAAAGTCTGCT 1685
 DB 457 LeuValIleProProValThrMetGlyPhePheValValLysAsnValAsnAlaLeuAla 476
 QY 1686 TGC 1688
 DB 477 Cys 477

RESULT 12

Q8T108 PRELIMINARY; PRT; 515 AA.
 ID 08T108
 AC 08T108;
 DT 01-JUN-2002 (TRENBLREL. 21, Created)
 DT 01-JUN-2002 (TRENBLREL. 21, Last sequence update)
 DT 01-JUN-2003 (TRENBLREL. 24, Last annotation update)
 DB Hepatase-like protein.

GN BMEPA.
 OS Bombyx mori (Silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
 OC Bombycidae; Bombyx.
 OK NCBI_TaxID=7091;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=p50; TISSUE=Posterior silk gland;
 RA Koike Y., Shimada T., Suzuki M.G., Mita K., Abe H., Maeda S.,
 RA Osogawa K., DeJong P.U.;
 RT "Genomic sequence of 320k containing a kettin orthologue on the Z
 chromosome in Bombyx mori";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=p50; TISSUE=Posterior silk gland;
 RA Koike Y., Shimada T., Suzuki M.G., Mita K., Abe H., Maeda S.,
 RA Osogawa K., DeJong P.U.;
 RT "Genomic sequence of 320k containing a kettin orthologue on the Z
 chromosome in Bombyx mori";
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB079860; BAB8191.1; -
 DR EMBL; AB090307; BAC10612.1; -
 DR InterPro; IPR005199; Glyco_hydro_79N.
 DR Pfam; PF03662; Glyco_hydro_79n; I.
 SQ SEQUENCE 515 AA; 59769 MW; FB8100ABE6EDDADB CRC64;
 Alignment Scores:
 Pred. No.: 8,366-48 Length: 515
 Score: 696.00 Matches: 183
 Percent Similarity: 50.96% Conservative: 83
 Best Local Similarity: 35.06% Mismatches: 182
 Query Match: 22.31% Indels: 74
 DB: Gaps: 18
 US-10-676-079-3 (1-1721) x Q8T108 (1-515)
 QY 198 CAGAGCCCGCTGACCTGATGACCCCTGCTTCGTCGATGACGACCACTG 257
 DB 42 GlnGluAspIleLysLeuIleSerGlnAspPheLeuSerPheGlyLysP---ThrIle 60
 QY 258 GCCAGGACCCGCGTTCCTCACTCTGCTGCTTCCAAAGCTTCTGATCTGGCCAGA 317
 DB 61 GluIleGluAsnTyrAsnArgIleAsnTyrSerAspThrArgLeuArgGluLeuAla 80
 QY 318 GCGTGTCTCTGCTGCTGATCTGATGATGATGATGATGATGATGATGATGATGAT 377
 DB 81 AlaLeuSerProIleArgLeuArgLeuGlyThrMetSerGluArgLeuIlePhe--- 99
 QY 378 CCCAAGAAAGATCAACCTTTGAAGAGAGAGATTAAGCAATCTCAAGTCAACAGAT 437
 DB 100 ---SerLysGluAsnIle-----ProIleSerCysHis 109
 QY 438 ATTGCAATATGATGATCCATCCCTGATGATGATGATGATGATGATGATGATGATGAT 497
 DB 110 AsnLysSerTyrLysSerTyrProLysSerLeuCysGln-----LeuIleGlySerPro 127
 QY 498 TACGAGAGCAATTGCTATCTCCGAGACATCAAGAAAGATTTGAAGAGAGACCTAC 557
 DB 128 Cys-----LysHisLysHisLysPheLeuProPhePheIleMet 140
 QY 558 TCAAGAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 617
 DB 141 ThrGlyAsnGluTrpAsnGlnIleAsnAspPheCysArgLysThrAsnLeuLysLeuLeu 160
 QY 618 TTGGCTAAATGCTTATTAAGACAGACAGATTTGACATGAGATTTATCTGAC 677
 DB 161 PheSerLeuAsnAlaMetLeuArgAsp---AsnHisGlyTrpAsnGluLysAsnAlaArg 179
 QY 678 TTGCTCTGAGACTGCTCTTCCAGGCGTATTAACATTTCTTGGGAATGAGCAATGAA 737
 DB 180 GluLeuIleGluPheSerLysHisLysGlnTyrAlaIleAspTrpGlnLeuGlyAsnGln 199

```

QY 738 CTTAAGGTTCTTAAAGAGGCTGATATTTTCAATGAGGCTCGAGCTTAAAGAGAT 797
DB 200 ProAenseRheglnh1sValRheAenGluSerValThrProGlnIleLeuAlaLeuAsp 219
QY 798 TAAATTCATGTCATTAATCTTAAAGAAAGTCCACCTTCAAAAATGCAAACTCTATGT 857
DB 220 PheGlnuYeuLeuArgLysLeuLeuAenIh1sValnIyLyrArgIh1sSerLeuIleValGly 239
QY 858 CTTGATGTTGGTCAAGCTT-----CGAAGAAAGAGGCTTAAGATGCTGAAGAGCTTC 908
DB 240 ProAseRhrThrArgProGlnProh1sArgProGluCyLeuYeuStyLyrMetIleGluPhe 259
QY 909 CTGAAGGCTGTGAGAGAGATGATTGATTCAGTTACATGCGATCATCATATTTGAATGA 968
DB 260 LeuGlnYeuGlnYeuGlnuAseRhrThrArgSerThrPheGlnIyLyrLeuAenSer 279
QY 969 CGAGAGCTGTCACAGAGAAAGTTTTCATTAACCTGATGATATTTGACATTTTATTCATCT 1028
DB 280 LysThrAlaLysLeuGlnuAseRhrThrArgProGlnuThrPheAseRhrLeu----- 296
QY 1029 GTGCAAAAAGTTTTCAGTGTGTTGAAGACACAGGCTGCGAAGAGTTC-----TGG 1082
DB 297 ArgGlnGlnIleGluThrMetGlnAenGlnThrLysLysStyLyrAsnIleProMetTrp 316
QY 1083 TTAGAGAAACAAGCTCTGCATATGAGGCGAGCGGCTTGTCTATCGACACCTTTCGA 1142
DB 317 LeuSerGlnThrSerSerStyLyrGlyGlyAlaProGlnYeuSerAenThrTyAla 336
QY 1143 GCTGGCTTATGTGGCTGATTAATTTGGGCTGTCAACCCGCAATGGGAATGAAAGTGTG 1202
DB 337 GlySerProLeuThrIleAseRhrLeuGlnYeuSerAlaLysThrAlaIleSerThrVal 356
QY 1203 ATGAGGCAATATTTCTTGGAGCAGAAATACCATTTAGTGTGAATTCATGATCTT 1262
DB 357 IleArgGlnSerPheIleGly---GlyTyLyrSerLeuValArgLysAenLeuLysPro 375
QY 1263 TTAACCTGATTATGGCTATCTTCTGTCAAGAAATGTGGCGACCAAGTGTTAATG 1322
DB 376 LeuProAseRhrTrpIleSerValLeuLysLysLeuValGlyAenLysValLeu--- 394
QY 1323 GCAGAGCTGCAAGTTCAAGAGAGAAAGCTTGCATATACCTTCTTCCACAAACACT 1382
DB 395 ---GlnAlaGlnCyAenCySerArgPheGlnArgLeuLysLysCysThrAenArg 413
QY 1383 GACAATCCAAAGTATTAAGAA---GAGATTTAATCTGTATGCCATAAAC----- 1430
DB 414 -----LysTyThrAenAseRhrSerAlaValIh1sLeuLysLysValAenLeuGln 430
QY 1431 -----CTCCATAAC-----GTC 1442
DB 431 MetAlaLysAlaArgPhePheLeuAenGlyThrAlaLeuIh1sGlyAseRhrLeuIle 450
QY 1443 ACCAAGTACTGGCGTTACCTTATCTTTTCATCAAGAAAGAGTAAATACCTTCTTA 1502
DB 451 H1sGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 463
QY 1503 AGACCTTTGGAGCTCATGATTACTTTCATTAATCTGTCCAACTCATGTGCTTAACCTTA 1562
DB 464 -----SerLysThrIleLeuLeuAenGlnLysLysLysLysLysLysLysLys 474
QY 1563 AAGATGTGATGATCAAACTTGGCACCTTTAAATGAAAAAAGCTTCGCGGACAGAAAGT 1622
DB 475 TyLyr-----GluSerAenLeuIh1sAenLeuArgProAenIleIh1sArgLysLys 492
QY 1623 TCACTGGGCTTGCAGCTTCTCATATAGTTTTTTTGTGATTAAGAAATGCCAAAGTTGCT 1682
DB 493 TyValSerLeuProLysSerIleGlyPheTrpValIleLysLysLysLysLysLys 512
QY 1683 GCTTGC 1688
DB 513 ValCys 514

```

```

RESULT 13
Q9SDA1 PRELIMINARY; PRT; 521 AA.
ID Q9SDA1;
AC Q9SDA1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN F13G24.30.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
ON NCBI_TaxID=5702;
RX [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Van Der Schueren J., Chuang Y.J., Voet M., Robben J.,
RA Volckaert G., Bancroft I., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: A133421; CAB62595.1; -.
DR PIR: T45608; T45608.
DR GO: GO:0004295; F:trypsin activity; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR005199; Glyco hydro 79N.
DR InterPro: IPR001254; Peptidase S1.
DR Pfam: PF03662; Glyco hydro 79n; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
KW Hypothetical protein.
SQ SEQUENCE 521 AA; 57831 MW; 07D8664A4B305CC2 CRC64;

Alignment Scores:
Score: No.: 4,85e-25 Length: 521
Score: 416.00 Matches: 154
Percent Similarity: 42.05% Conservative: 68
Best Local Similarity: 29.17% Mismatches: 184
Query Match: 13.34% Indels: 122
DB: Gaps: 24

US-10-676-079-3 (1-1721) x Q9SDA1 (1-521)
QY 285 CTGGGTTCTCCAAAGCTTGATCTTGGCCAGAGGCTTCTCTGCTGATCGATGATTT 344
DB 55 LeuThrArgProLeuLeuThrLysAlaIleLysAlaPheLysProLeuLysArgLysIle 74
QY 345 GGTGGCAACAAGCAAGCTTCTTAATTTTCGATCCCAAGAAAGATCAACCTTGAAGAG 404
DB 75 GlyGlySerLeuGlnAenArgLysValIleLysArgValGlyAenLeuLysThr----- 91
QY 405 AGAAGTTACTGCGAATCTCAAGTCAACAGATATTGCAAAATATGATTCATCCCTCTCT 464
DB 92 -----ProCyAArg----- 94
QY 465 GATGTGAGAGAGATTAACGTTGGAATGAGCCCTTACAGAGCAAGTAATGTCTATCCAGAA 524
DB 95 -----ProPheGlnLysMet----- 99
QY 525 CACTACCCAGAAAAAGTTCAAGAAAGC-----ACCTATCAAGAAAGCTCTGTA--- 572
DB 100 -----AenSerGlyLeuPheGlyPheSerLysLysLysLysLysLysLysLys 112
QY 573 -----GATGTGCTATCACTTTTTCGAAATGTGCGAGGAGCTGAGCTGATCTTT 620
DB 113 MetLysArgTrpAseRhrLeuAenSerPheLeuThrAlaThrGlyAlaValAlaThrPhe 132
QY 621 GGCCTAAATGCGTTATTAAGAACAGCAGATTTGCAAG-----TGG 659
DB 133 GlyLeuAenAlaLeuArgLysLysLysLysLysLysLysLysLysLysLysLysLysLys 152
QY 660 AACAGTTCTTAATGCTCAAGTTGCTCTGAGACTACTGCTTCCAAAGGAGTAAACATT--- 716

```

```

Db 153 Asph1e1leantThrGlnasphPheleuanTyrtThValserLygLyTyrtVal1leasp 172
Qy 717 TTTTGGAACTTGGCAAGTGAACCTTAACGCTTCTTAAGAAGGCTGATATTTTCATCAT 776
Db 173 SerTrpGlnPheGlyAsnGlnLeuSerGly-----SerGlyValGlyAlaSerValser 190
Qy 777 GGGTCGACGATTAGAGAGATATATATTCATTAACCTTAAGAAGGCAAGCTTC 836
Db 191 AlaGlnLeuTyrtGlyLeuAspLeuIleValLeuLyAspValIleAsnLyse--ValTyrt 209
Qy 837 AAAAATGCAAACTGTATGCTTCGTATGCTTGGTCAAGCTT-----CGAAGA 881
Db 210 LysAsnSerTrpLeuIleTyrtProIleLeuValAlaProGlyGlyPheTyrtGlnGln 229
Qy 882 AAGACGGCTAAGATGCTGAAGAGCTTCTCGAAGCGCTGGTGGAGAAAGTATTCATTCAGTT 941
Db 230 TrpTyrtThLyseLeuGlnIle-----SerGlyProSerValValAlaAspValVal 246
Qy 942 ACATGGCACTCACTATATTGAATGAGCGAGCT-----GCTACGAGGAGATTTTCTA 995
Db 247 ThrIleSh1e1leTyrtAsnLeuGlySerGlyAsnAspProIleAlaLeuValLyseIleMet 266
Qy 996 AACCTGTATGTATGACATTTTATTCATCTGTGCAAAAAGTTTTC-----1043
Db 267 AspProSer-----TyrtLeuSerGlnValSerLyThrPheLyAspValAsn 282
Qy 1044 CAGGTGTGTAGAGACCAAGGCTGGCAAGAGCTGTGTTGGAGAAACAGCTCTGCA 1103
Db 283 GlnThrIleGlnGlnIleGlyProTrpAlaSerProTrpValGlyGlnSerGlyGlyAla 302
Qy 1104 TATGAGAGCGAGCGCCCTTGTATCCGACCTTGTGACGCTGGCTTATGNGGCTGAT 1163
Db 303 TyrtAsnSerGlyGlyAlaArgHisValSerAspThrPheIleAspSerPheTrpTyrtLeuAsp 322
Qy 1164 AAATTTGGGCTTCAGCCCGAATGGGAATAGAAGTGTGATAGGCAAGTATTTCTTGA 1223
Db 323 GlnLeuGlyMetSerAlaArgHisAsnThrLyseValTyrtCyAsArgIleThrLeuValGly 342
Qy 1224 GCAGAAACTACCATTTAGTGATGAAGA---AACTTCGATCCTTTACCTGATATTGGCTA 1280
Db 343 ---GlyPheTyrtGlyLeuLeuGlnLyseGlyThrPheValProAspProAspTyrtLyser 361
Qy 1281 TCTCTTGTGTAGAAATTTGGTGGGCAAGGCTTATATGGAAGCCTGCAAGCTTCA 1340
Db 362 AlaLeuLeuTrpHisArgLeuMetGlyLyseGlyValLeuAlaValGlnThrAspGlyPro 381
Qy 1341 AAGAGAAAGAGCTTCGATATACCTTCATTTGCAAAACATGACATCCAAAGTATATA 1400
Db 382 Pro-----GlnLeuArgValTyrtAlaHisCySerLyse-----Gly 393
Qy 1401 GAAGAGATTTAAGCTGTATGCCATTAACCTTCATACGTCACCAAGTACTTG-----1454
Db 394 ArgAlaGlyValThrLeuLeuLeuLeuLeuLeuLeuSerAsnGlnSerAspPheThrValSer 413
Qy 1454 -----1454
Db 414 ValSerAsnGlyIleAsnValValLeuAsnAlaGlnSerArgLyseLyseLyseLeu 433
Qy 1455 ---CGTTACCTTATCTTTTCT-----AACAGCAAGTGGATTAATACCTTCTA 1502
Db 434 AspThrLeuLeuArgProPheSerTrpIleGlySerLyseAlaSerAspGlyTyrtLeuAsn 453
Qy 1503 AAGACT-----TTGGACCTCATGGA-----TACTTTCCAAATCTGTCAACTC 1547
Db 454 ArgGlnGlnTyrtHisLeuThrProGlnValAsnGlyValLeuArgSerLyseThrMetValLeu 473
Qy 1548 AATGGCTAAGCTTAAGATGATGATGATCAAACTTGGCAAGCTTATGAGAAAACCT 1607
Db 474 AsnGlyLyseSerLeuLyseProThrAlaThrGlyAspIleProSerLeu---GlnProVal 492
Qy 1608 CTCGCGGCA---GGAATTCACCTGGGCTTGCAGGCTTTCTCATATAGTTTGTGANA 1664

```

```

Db 493 LeuArgSerValAsnSerProLeuAsnValLeuProLeuSerMetSerPheIleValLeu 512
Qy 1665 AGAAATGCCAAAGTTCGCTTCG 1688
Db 513 ProAsnPheAspAlaSerAlaCys 520

RESULT 14
ID 09FF10 PRELIMINARY; PRT; 543 AA.
AC 09FF10;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Similarity to heparanase.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT features of the 1.6 Mb regions covered by twenty physically assigned
RT pl clones."
RL DNA Res. 4:215-230(1997).
DR EMBL; AB005249; BAB09947.1;
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR005199; Glyco_hydro_79N.
DR InterPro; IPR001254; Peptidase_S1.
DR Pfam; PF03662; Glyco_hydro_79n; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
SQ SEQUENCE 543 AA; 60250 MW; 0FA2248948282FF6 CRC64;

Alignment Scores:
Pred. No.: 4.92e-25 Length: 543
Score: 416.00 Matches: 154
Percent Similarity: 42.05% Conservative: 68
Best Local Similarity: 29.17% Mismatches: 184
Query Match: 13.34% Indels: 122
Dels: 10 Gaps: 24

US-10-676-079-3 (1-1721) x 09FF10 (1-543)
Qy 285 CTGGGTTCTCCAAAGCTTCGTACCTTGGCGAGAGGCTGTCTCTGCGTACCTGAGTTT 344
Db 77 LeuThrArgProLeuLeuThrLyseAlaIleLyseAlaPheLyseProLeuArgIleArgIle 96
Qy 345 GGTGGCAACCAAGACAGCTTCTTAATTTGATCCCAAGAAAGATCAACCTTGAAGAG 404
Db 97 GlyLyseSerLeuGlnAspGlnValIleTyrtAspValGlyAsnLeuLyseThr-----113
Qy 405 AGAAGTTACTGGCAATCTCAAGTCAACCAAGATATTGGCAATATGATTCATCCCTCT 464
Db 114 -----ProCyAsArg-----116
Qy 465 GATGTGAGAGAAAGTTACGGTTGGAATGGCCCTACCAAGAGCAATGCTACTCCGAGAA 524
Db 117 -----ProPheGlnLyseMet-----121
Qy 525 CACTACAGAAAAAGTTCAAGACAGC-----ACCTACGACAGAGCTGTGA---572
Db 122 -----AsnSerGlyLeuPheGlyPheSerLyseGlyCysLeuHis 134
Qy 573 -----GATGCTATACACTTTTGGCAAGCTGTACAGAGCTGATCTT 620
Db 135 MetLyAsArgTrpAspGlnLeuAsnSerPheLeuThrAlaThrGlyAlaValThrPhe 154
Qy 621 GGCCTTAATGCGTTATTAAGAAACAGACAGATTGCG-----TGG 659

```

[illegible]

Db	496	Asn61glyssrleu5yrprothralathrc1ylnap1leproserleu---glu1proval	514
Qy	1608	CTCCGGCCCA---GGAAGTTACATCGGGCTTGCCACAGCTTTCATATAGTTTTTGTGATA	1668
Db	515	Leu4rsgersValaanserProleuaasval1leuProleuserMetserPhe1leValleu	534
Qy	1665	AGAAATGCGCAAGTTGCTGCTGC	1688
Db	535	ProasnheapalaseralaCys	542
RESULT 15			
Q89F99	PRELIMINARY;	PRT;	559 AA.
ID	Q89F99;		
AC	Q89F99;		
DT	01-JUN-2003 (TREMBlrel. 24, Created)		
DT	01-JUN-2003 (TREMBlrel. 24, Last sequence update)		
DT	01-JUN-2003 (TREMBlrel. 24, Last annotation update)		
DE	Blr6802 protein.		
GN	Blr6802.		
OS	Bradyrhizobium japonicum.		
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;		
OC	Bradyrhizobiaceae; Bradyrhizobium.		
OX	NCBI_TaxID=375;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=USDA 110;		
RC	MEDLINE=22484998; PubMed=12597275;		
RX	Kaneko T., Nakamura Y., Sato S., Minamitawa K., Uchiumi T.,		
RA	Sasamoto S., Watanabe A., Ideawa K., Iiguchi M., Kawashima K.,		
RA	Kohara M., Matsumoto M., Shimo S., Teurtocka H., Wada T., Yamada M.,		
RA	Tabata S.;		
RT	"Complete genomic sequence of nitrogen-fixing symbiotic bacterium		
RT	Bradyrhizobium japonicum USDA110."		
RL	DNA Ref. 9:189-197(2002).		
DR	EMBL; AP005959; BACS2067.1; -		
KW	Complete proteome.		
SQ	SEQUENCE 559 AA; 59948 MW; F3B423547MA30170 CRC64;		
Alignment Scores:			
	Pred. No.:	2.94e-24	Length: 559
	Score:	406.50	Matches: 164
	Percent Similarity:	41.57%	Conservative: 80
	Best Local Similarity:	27.94%	Mismatch: 211
	Query Match:	13.03%	Indels: 132
	DB:	16	Gaps: 28
US-10-676-079-3 (1-1721) x Q89F99 (1-559)			
Qy	62	GATGCTGTGGCGCTCGAAGCCTGGCGCTGCCCGCGCTGATGCTGCTGCTCGGGGC	121
Db	60	AspAlaAlaArgVal1ThrGlyLeuAlaLeuLeuAlaAlaGlyThrSerCysAlaAlaHis	79
Qy	122	GCTGGG-----TCCCGCTCCCGCTGGCGCCCTGGCGCCGACCTGGGCAAGACAGAGCT	175
Db	80	AlaGlyThrValSer-IleAlaProAlaGlyLeu-----	90
Qy	176	CGTGGACCTGGACTTCTTACCCAGAGACCGCTGCACCTGTGAGCCCTCGTTCCTGC	235
Db	91	-----ArgAlaIleGlyThrIleAspProAlaGlnSer	102
Qy	236	CGTACCACTTGACGCCCAACTGGCCAGGACCCGGGTTCTC-----	278
Db	102	ITyAsnIleGlnMetValGluVal1ThrGlyGlyArgPheThrPlyPProTyProGlnAl	122
Qy	279	-----ATCCCTCGGGCTTC	292
Db	122	ameArAla1ATpAlaAspLysAspArgTySerTyArgProProIleAspLeuGlyAs	142
Qy	293	TCCAAAGCTTGTGATCTTGCGCAAGGCTTGTCTCTGCGTACCTGAGAGTTGTGGAC	352
Db	142	nThArgLeuArgGlyLeuAlaValAlaLeuSerProAlaTyLeuArgValSerGlyTh	162
Qy	353	CAAGACAGACTTCTTAATTTTCGATCCCAAGAGATCAACTTTGAAGAGAGATT	412


```

Db      162 rTrrAlaAsnAlaThrPhePheAlaAspThrGluSerAlaPro----- 176
Qy      413 CTGGCAATCTCAACGACACGAGATATTGGCAAAATATGATTCATCCCTCTGATGTGA 472
Db      177 -----AlaAlaProProGlyPheAs 184
Qy      473 GGAGAAGTTA---CGGTGAATGGCCCTACAGAGACAAATTGTACTCCAGAACACTA 529
Db      184 nAlaValLeuGlyAArgAlaGlnTrp----- 192
Qy      530 CCAGAAAAGTTCAAGAACAGACCTTCAAGAACTGTAGATGTGCTATACACTT 589
Db      193 -----ArgGlyValAlaAsp-----Ph 198
Qy      590 TGCCAACTGCTCAGAGACTGATGATCTTTGGCCTA-----AATGCTTATTAAAGAC 643
Db      198 eAlaArGAlaThrGlyAlaGlnIleValThrSerLeuAlaValSerProGlySerArgAs 218
Qy      644 AGCAGATTTCAGAGAACAGTTCTAATGCTCAGTTGCTCTGACTACTGCTTCCAA 703
Db      218 pAlaAspGlyLeuTrpArgProAspGlnAlaGlnArgLeuIleAspTrpThrArgSerIle 238
Qy      704 GGGGTATTAACATT---TCTTGGAACTAGGCAATGAACTTACAGTTTCCTTAAGAAAGC 760
Db      238 nGlyGlyHisIleAlaAlaGluPheMetAsnGluProThrLeu----- 253
Qy      761 TGATATTTTCATCATATGGGTCCGAGTTAGAGAAAGT-----TATATTCAA----- 806
Db      254 ----AlaAlaThrAsnGlyAlaProProGlyTrpAspAlaIyValIyValArgAspVa 272
Qy      807 -----TTGCATAAACTTCTAAGAAAGTCACCTTCAAAAATGCAAAACTCATGGTCC 859
Db      272 lGlyIlePheHisGluTrpMetLeuArgAlaAla---ProGlnThrLeuIleValGlyPr 291
Qy      860 T---GATGTGGTCAAGCTCGAAGAAAGACGGCTAAGATGCTGAAGAGC-----TTCCT 910
Db      291 oGlySerValGlyAspSerAlaSerThrGlyAlaSerGlyLeuArgThrArgAspLeuIle 311
Qy      911 GAAGCTGTGTGAGAAAGTATGATTCAGTTACATGCGCATCACTAC---TATTTGAATGG 967
Db      311 nAlaAlaSerGlyAlaGlyValAspArgPheSerTrpHisIleTyAsnThrIleSerPr 331
Qy      968 AGGACCTGCTACACAGGAAGATTTTCTAACCCTGATGATTGAGACATTTTATTTCATC 1027
Db      331 oArgCysGlyGlyArgAspGlnProAlaGlyAla---LeuSerGluAlaTrpLeuAlaAr 350
Qy      1028 TGTGCAAAAAGTTTCCAGGTGGTGAAGACACCAAG-----CCTGGCAAGAA 1075
Db      350 gThrAspAlaAlaLeuSerThrTrpLysSerLeuArgAspGluPheAlaProAspLysPr 370
Qy      1076 GGTCTGGTTAGAGAAACAGCTCTGCAATATGAGAGCGGAGCGCCCTTCTATCCGACAC 1135
Db      370 oIleTrpLeuThrGluThrAlaAspAlaAlaCysGlyGly---AsnArgTrpAspLysPr 389
Qy      1136 CTTTGACAGTGGCTTATATGCTGATTAATTGGGCGCTGACGCCGAATGGGAATAGA 1195
Db      389 rPheLeuAspThrPheArgTrpLeuAspGlnLeuGlyArgLeuAlaArgAlaGlyValG1 409
Qy      1196 AGTGTGATGAGGCAAGTATTTCTTGGAGCAGGAAACTACCATTTAGATGAA---AA 1252
Db      409 nValValMetHisAsnThr---LeuAlaAlaSerAspTrpGlyLeuLeuAspGlnLysPr 428
Qy      1253 CTTGCATCTTTACTGATTAATGCTATCTCTTCTTCAAGAAATGGTGGGACCA 1312
Db      428 rPheArgProArgProAsnTrpArgIyAlaLeuLeuTrpArgArgLeuMetGlyThrTrh 448
Qy      1313 GGTGTTAATGAGCAAGGTTCAAGAGGAGGAAGAGGAGTTCAGATACTTACCTCATG 1372
Db      448 rValLeuAspAlaGly-----AlaAlaMetAlaProGlyLeuHisValIyAlaHisCy 466
Qy      1373 CACAAACACTGCAATCCAAAGGATATTAAGAGAGATTAACTGTATGCAATAACCT 1432

```

```

Db      466 s-----HisProSer---LysArgGlyAlaValAlaThrValLeuAlaIleAsnI1 481
Qy      1433 CCATTAAGCTCAC---AGTACTTGCGGTACCCGATACCTTTTCTAACAAGACAGTGA 1489
Db      481 eSerArgSerThrAlaArgThrIleValLeuProLeuPro-----AlaG1 496
Qy      1490 TAAATACCTTTAAGACCTTTGGGACCTCATGATTACTTTCAAAATCTGTCCAACCTGA 1549
Db      496 uArgTrpThrLeuGlnAlaAlaArgLeuGlnGly-----AlaThrValGlnLeuAs 513
Qy      1550 TGGTCTAATCTTAAAGATGTGATGATCAAACTTGCACCTTTAATGAAAAACCTCT 1609
Db      513 nGlyLysThrLeuAlaLeuThrAspGluAspGlyLeuProProLeuAlaGlyArgAlaI1 533
Qy      1610 CCGGCGAAGAAAGTTCACTGGGCTTGCACGCTTCTCATATAGTTTGTGATPAAGAA 1669
Db      533 eGluAlaGly---AlaValGlnLeuAlaSerGlnThrIleThrPheLeuAlaIleProG1 552
Qy      1670 TGCCAAGTTGCTGCTTGC 1688
Db      552 yAlaAlaAsnSerAlaCys 558

```

Search completed: August 28, 2004, 07:51:40
Job time : 248.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: August 28, 2004, 07:42:36 ; Search time 193 Seconds
(without alignments)
5039.008 Million cell updates/sec

Title: US-10-676-079-3
Perfect score: 3119
Sequence: 1 ctgagcttcgactctccg.....atactagctcctgactc 1721

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 28547505 residues
Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_n2p.model -DEV=xlP
-O=/cgn2.1/USPTO_epool_p/US10676079/runat_28082004_084224_23795/app_query.fasta_1.1863
-DB=A.GeneSeq_29Jan04 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPTCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10676079 @CGN 1 1 263 @runat_28082004_084224_23795 -NCPU=6 -ICPU=3
-NO MAP -LANG=QUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: GeneSeq_29Jan04:*
2: GeneSeqp1980s:*
3: GeneSeqp1990s:*
4: GeneSeqp2000s:*
5: GeneSeqp2001s:*
6: GeneSeqp2002s:*
7: GeneSeqp2003s:*
8: GeneSeqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2949	94.5	592	2	AA023346 A human h
2	2949	94.5	592	3	AA080850 Amino aci
3	2904	93.1	588	2	AA030124 A human p
4	2842	91.1	543	2	AA023345 A human h
5	2842	91.1	543	3	AA057590 Human hep
6	2842	91.1	543	3	AA080849 Amino aci
7	2842	91.1	543	3	AA052990 Human hep
8	2842	91.1	543	4	AA076635 Human hep
9	2842	91.1	543	5	AA070813 Human hep
10	2838	91.0	543	2	AA017082 Human hep

11	2838	91.0	543	4	AA086206 Human hep
12	2838	91.0	543	7	ADD18950 Human dis
13	2826	90.6	543	4	AA088361 Human mem
14	2817	90.3	545	6	ABP56822 Human hep
15	2817	90.3	545	7	AD016012 G-coupled
16	2772	88.9	532	2	AA017083 Human pre
17	2764	88.6	530	2	AA034173 Chicken s
18	2673.5	85.7	527	5	AB007815 A murine
19	2146	68.8	535	3	AB008851 Mouse hep
20	2146	68.8	535	5	AB007811 Rat hep
21	2123	68.1	536	5	AB007812 Rat hep
22	1645.5	52.8	523	3	AB007814 Chicken h
23	1614	51.7	380	2	AA017085 Human hep
24	1602	51.4	380	2	AA017084 Mouse hep
25	1154.5	37.0	592	4	AA07632 Human hep
26	1154.5	37.0	592	4	AA076424 Human hep
27	1148.5	36.8	592	4	AA081062 Human hep
28	1147.5	36.8	592	4	AA085215 Hepatarnas
29	1142.5	35.7	582	5	AA018326 Human hep
30	1112.5	35.7	538	4	AA097633 Human hep
31	1106.5	35.5	528	5	AA018327 Human hep
32	936.5	30.0	534	4	AA085216 Hepatarnas
33	936.5	30.0	534	5	AB069310 Human pol
34	936.5	30.0	534	5	AA050337 Human pre
35	927.5	29.7	492	4	AA084664 Amino aci
36	897.5	28.8	480	4	AA097634 Novel hum
37	897.5	28.8	480	4	AA007418 Human hep
38	897.5	28.8	480	4	AA085217 Hepatarnas
39	892.5	28.6	470	5	AA018328 Human hep
40	891.5	28.6	439	4	AA007423 Human hep
41	788	25.3	331	5	AA050383 Human hep
42	663	21.3	488	4	AA031469 Amino aci
43	645	20.7	488	4	AA031470 Amino aci
44	642	20.6	488	4	AA031472 Amino aci
45	622	19.9	488	4	AA031471 Amino aci

ALIGNMENTS

RESULT 1
AA023346
ID AA023346 standard; protein; 592 AA.
XX
AC AA023346;
XX
DT 09-JUL-1999 (first entry)
XX
DE A human heparanase protein.
XX
KW Heparanase; hpa; modulator; heparin-binding growth factor;
KW cellular response; cytokine; cell interaction; plasma lipoprotein;
KW cellular susceptibility; infection; disintegration;
KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease; neuritis;
KW plasma heparin; micrometastasis; autoimmune lesion; renal failure.
XX
OS Homo sapiens.
XX
PN WO9911798-A1.
XX
PD 11-MAR-1999.
XX
PF 31-AUG-1998; 98MO-US017954.
XX
PR 02-SEP-1997; 97US-00922170.
XX
PR 02-JUL-1998; 98US-00109386.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
PA (HADA-) HADAST MEDICAL RES SERVICES & DEV.
PA (FRIE/) FRIEDMAN M M.
XX
PI Pecker I, Vlodavsky I, Feinstein E;
XX

DR WPI; 1999-302255/25.
 DR N-PSDB; AAX35650.
 XX
 PT New human polynucleotide useful for treating angiogenesis, restenosis,
 and inflammation.
 PS
 XX Claim 6; Page 65-66; 63pp; English.
 CC The specification describes a polypeptide having heparanase (hpa)
 CC activity. The recombinant protein is used as a modulator of heparin-
 CC binding growth factors, cellular responses to heparin-binding growth
 CC factors and cytokines, cell interaction with plasma lipoproteins,
 CC cellular susceptibility to viral, protozoal and bacterial infections or
 CC disintegration of neurodegenerative plaques. Heparanase may be useful for
 CC conditions such as wound healing, angiogenesis, restenosis,
 CC atherosclerosis, inflammation, neurodegenerative diseases, and viral
 CC infections. Mammalian heparanase can be used to neutralize plasma
 CC heparin, and anti-heparanase antibodies may be applied for
 CC immunodetection and diagnosis of micrometastases, autoimmune lesions, and
 CC renal failure in biopsy specimens, plasma samples, and body fluids. The
 CC present sequence represents human heparanase

SO Sequence 592 AA;

Alignment Scores:

Pred. No.:	2,98e-283	Length:	592
Score:	2949.00	Matches:	562
Percent Similarity:	99.82%	Conservative:	0
Best Local Similarity:	99.82%	Mismatches:	1
Query Match:	94.55%	Indels:	0
DB:	2	Gaps:	0

US-10-676-079-3 (1-1721) x AAY02346 (1-592)

QY 3 AGAGCTTTCAGCTCTCCGCGCGAGCTGCGGGGGGAGCAGCCAGGTGAGCCCAAG 62
 DB 30 ArgAlaLeuaspSerProLeuArgGlySerTrpArgIylGlnProGlyGlnProLys 49
 QY 63 ATGCTGCTGCGCTCGAAGCTGCGCTGCGCGCGCGCTGATGCTGCTCTTGAGCGG 122
 DB 50 MetLeuLeuArgSerIylsProAlaLeuProProLeuMetLeuLeuGlyPro 69
 QY 123 CTGGGTCCTCTCTCCCTGCGCGCTGCGCGCGCTGCGCGAGCAGCAGAGCTGTGAC 182
 DB 70 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaGlnAspValAsp 89
 QY 183 CTGAGCTTCTTCAACCCAGAGCGCGCTGACCTGTGAGCCCTGCTCCGTCCGTACC 242
 DB 90 LeuAspPhePheThrGlnGlnProLeuHleuValSerProSerPheLeuSerValThr 109
 QY 243 ATTGACGCCAAGCTGCGCAGCAGCCCGGTTCTCATCTCTGAGGTTCTCCAAAGCTT 302
 DB 110 IleAspAlaAsnLeuAlaThrAspProArgPheLeuIleLeuGlySerProLysLeu 129
 QY 303 CGTACCTGGCGCAGAGCTGTCTCTGCGTATCTGAGGTTGGTGACCAAGACGAC 362
 DB 130 ArgThrLeuAlaArgGlyLeuSerProAlaIylLeuArgPheGlyIylThrIylsThrAsp 149
 QY 363 TTCTCTATTTTCGATCCCAAGAGGAATCAACCTTGAAGAGAGATTACTGGCAATCT 422
 DB 150 PheLeuIlePheAspProLysIylsGlnSerThrPheGlnGlnIylsGlnSer 169
 QY 423 CAAGTCAACAGATATTTGCAATATGATTCATCCCTCTGATGTGAGAGAGATTA 482
 DB 170 GlnValAsnGlnAspIleCyIylsIylsGlnSerIleProAspValGlnGlnIylsLeu 189
 QY 483 CGGTGGATGCGCTTACAGAGAGCAATGCTACTCCGAGAACATACAGAAAAGTTC 542
 DB 190 ArgLeuGlnIylsProIylsGlnGlnIylsLeuLeuArgGlnIylsIylsIylsPhe 209
 QY 543 AAGAAACGACCTTACTAAGAGCTGTAGATGTGTATACATTTTGGAACTGTCTCA 602
 DB 210 LysAsnSerThrIylsSerArgSerSerValAspValLeuIylsThrPheAlaAsnCySer 229

QY 603 GCACTGACCTGATCTTTGGCCCTAAATGCGTTATTAAGAACAGCAGATTGCGAAG 662
 DB 230 GlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaPheLeuGlnTrpAsn 249
 QY 663 AGTTCTAATGCTCAGTTGCTCTGAGACTACCTGCTCTCCAGGGGTATTAACATTTCTGG 722
 DB 250 SerSerAsnAlaGlnLeuLeuLeuAspIylsCySerSerIylsGlyIylsAsnIleSerTrp 269
 QY 723 GAATGAGGCAATGAACCTTAACAGTTTCTTAAGAGCGCTGATATTTTCATATGAGGTG 782
 DB 270 GlnLeuGlyAsnGlnProAsnSerPheLeuLysValAspIlePheIleAsnGlySer 289
 QY 783 CAGTTGAGAGAGATTAATTAATTCATTAATGATTAACCTTAAGAAAGTCCACTTCAAAAT 842
 DB 290 GlnLeuGlyIylsLeuPylsIylsGlnLeuHleuHleuLysLeuLeuArgIylsSerThrPheLysAsn 309
 QY 843 GCAAAACCTATATGCTCTGATGTTGGTCAGCTCCGAGAAAGACCGCTTAAGATGCTGAAG 902
 DB 310 AlaLysLeuIylsIylsProAspValGlyGlnProArgArgIylsThrAlaLysMetLeuLys 329
 QY 903 AGCTTCCTGAGGCTGCTGAGAGAGATTAATTCAGTTACATGAGCAGTACTATTTTG 962
 DB 330 SerPheLeuLysAlaGlyIylsGlnValIleAspSerValThrTrpHisIylsIylsLeu 349
 QY 963 AATGACGAGCTGCTACAGAGAGATTTTCTAAACCTGATGATTTGACATTTTATTT 1022
 DB 350 AsnGlyArgThrAlaThrArgGlnAspPheLeuAsnProAspValLeuAspIlePheIle 369
 QY 1023 TCATCTGTGCAAAAGATTTTCCAGGTGTGTGAGAGACCAAGGCTGTGCAAGAGCTGTG 1082
 DB 370 SerSerValGlnLysValPheGlnValIylsGlnSerThrArgProIylsIylsValTrp 389
 QY 1083 TTAGAGAAACAAGCTCTGATATGAGAGCGAGCGCCCTGCTATCCGACACTTTGCA 1142
 DB 390 LeuGlyGlnThrSerSerAlaIylsIylsGlyIylsAlaProLeuLeuSerAspThrPheAla 409
 QY 1143 GCTGGCTTATATGCTGAGATAAATTTGGCTGTGACCCGAAATGGGAAATGAAGTGTG 1202
 DB 410 AlaGlyPheMetTrpLeuAspLysLeuGlyLeuSerAlaArgMetGlyIleGlnVal 429
 QY 1203 ATGAGGCAATATTTCTTTGAGAGAGAACTACATTTAGTGATGAATAAACTTCGATCCT 1262
 DB 430 MetArgGlnValPhePheGlyAlaGlyAsnIylsIylsLeuValAspGlnAsnPheAspPro 449
 QY 1263 TTACTGATTAATGCTATCTCTCTGTTCAAGAAATTTGTGGGCAACAGGTGTTAATG 1322
 DB 450 LeuProAspTrpTrpLeuSerLeuLeuPheLysIylsLeuValGlyIylsIylsValLeuMet 469
 QY 1323 GCAAGGCTGCAAGGTTCAAGAGAGAAAGCTTCAAGTATACCTTGATTCACAAACACT 1382
 DB 470 AlaSerValGlnGlySerIylsAspArgLysLeuArgValIylsLeuHleuHleuCySerThrAsnThr 489
 QY 1383 GACATTCGAAGATTAATAAGAGAGATTTTAATCTGTATAGCCATAAACCCTCCATAACGTC 1442
 DB 490 AspAsnProArgIylsIylsGlnGlyAspLeuThrIleuIylsAlaIleAsnLeuHleuAsnVal 509
 QY 1443 ACCAAGTACTTGGGTTACCTATCTTTCTTAACAACAAGTGAATTAATCTTTCTA 1502
 DB 510 ThrIylsIylsLeuArgLeuProIylsProPheSerAsnLysGlnValAspIylsIylsLeu 529
 QY 1503 AGACCTTTGGGACCTGATGATTAATTTCCAAATCTGTCCAACTCAATGATCTTAATCTCTA 1562
 DB 530 ArgProLeuGlyProHleuGlyLeuLeuSerLysSerValGlnLeuAsnGlyLeuThrLeu 549
 QY 1563 AAGATGTGATGATTAACCTTTGCCACCTTTTAATGAGAAAACCTTCGCGGCAAGAAAT 1622
 DB 550 LysMetValAspAspGlnThrLeuProProLeuMetGlnLysProLeuArgProGlySer 569
 QY 1623 TCATGAGGCTGTGACCTTCTCATATAGTTTTTTTGTATAGAAATGCAAGATGTGCT 1682
 DB 570 SerLeuGlyLeuProAlaPheSerIylsSerPhePheValIleArgAsnAlaLysValAla 589

QY 1683 GCTTCATC 1691
 DB 590 AlaGySile 592

RESULT 2
 AAB08850
 ID AAB08850 standard; protein; 592 AA.

XX AAB08850;
 AC
 XX 15-JAN-2001 (first entry)

DE Amino acid sequence of a human heparanase polypeptide.
 KW Human; heparanase; gene therapy; tumour; inflammation; neurodegenerative plaque;
 KW heparin-binding growth factor; cytokine; neurodegenerative disease;
 KW wound healing; infection; burn; angiogenesis; restenosis;
 KW atherosclerosis; inflammation; neurodegenerative disease;
 KW Gerstmann-Strausler Syndrome; Creutzfeldt-Jakob disease.

OS Homo sapiens.
 XX WO200052178-A1.
 XX 08-SEP-2000.
 XX 14-FEB-2000; 2000MO-US003542.
 XX 01-MAR-1999; 99US-00258892.

PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 PA (FRIE/) FRIEDMAN M M.
 XX Pecker I, Vlodevsky I, Feinstein E,
 DR WPI; 2000-579289/54.
 DR N-PSDB; AAA75053.

PT New polynucleotides encoding a polypeptide having heparanase activity,
 PT useful in wound healing and in gene therapy, particularly in treating
 PT tumor, inflammation, autoimmunity, neurodegenerative diseases.

PS Claim 22; Page 122-123; 152pp; English.

XX The present sequence represents a human protein with heparanase catalytic
 CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
 CC particularly in treating tumour, inflammation or autoimmunity.
 CC Particularly, the polynucleotide is useful in modulating the
 CC bioavailability of heparin-binding growth factors, cellular responses to
 CC heparin-binding growth factors (e.g. bFGF) and cytokines (e.g.
 CC interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular
 CC susceptibility to certain viral and some bacterial and protozoa
 CC infections, or disintegration of neurodegenerative plaques. The
 CC polynucleotide is also useful in wound healing (e.g. thermal, chemical or
 CC radiation burns), and in the treatment of angiogenesis, restenosis,
 CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
 CC Strausler Syndrome or Creutzfeldt-Jakob disease), and some viral,
 CC bacterial or protozoa infections

SQ Sequence 592 AA;

Alignment Scores:
 Pred. No.: 2,98e-283 Length: 592
 Score: 2949.00 Matches: 562
 Percent Similarity: 99.82% Conservative: 0
 Best Local Similarity: 99.82% Mismatches: 1
 Query Match: 94.55% Indels: 0
 DB: 3 Gaps: 0

US-10-676-079-3 (1-1721) x AAB08850 (1-592)

QY 3 AGAGTTTGACTCTCGCTGCGCGGCACTGCGCGGGGAGACAGCCAGGTAGCCCAAG 62

DB 30 ArgAlaLeuApsSerProLeuArgGlySerTrpArgGlyGluGlnProGlyGluProIys 49
 QY 63 ATGCTGCTGGCGCTGCAAAACCTGCGCGCGCGCGCGGATGCTGCTGCTGCGGCGCG 122
 DB 50 MetLeuLeuArySerLysProAlaLeuProProProLeuMetLeuLeuLeuLysPro 69
 QY 123 CTGGGTCCTCTCTCTCTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182
 DB LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaGlnAspValValAsp 89
 QY 183 CTGACCTTCTTCACCCAGAGCGCTGCACTGCTGACCCCTGCTGCTGCTGCTGCTCAC 242
 DB LeuAepPhePheThrGlnGluProLeuHisLeuValSerProSerPheLeuSerValThr 109
 QY 243 ATTGACGCCAACCTGCG 302
 DB 110 IleAepAlaAenLeuAlaThrAspProArgPheLeuIleLeuLeuGlySerProLysLeu 129
 QY 303 CGTACCTTGGCCAGAGGCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 362
 DB ArgThrLeuAlaArgGlyLeuSerProAlaTyrLeuAryPheGlyGlyThrLysThrAsp 149
 QY 363 TTCTTAATTTTCGATCCCAAGAGGAATCAACTTTGAAGAAGAGATTACTGCGCAATCT 422
 DB PheLeuIlePheAepProLysGlyLeuSerThrPheGlnGluArgSerTyrTrpGlnSer 169
 QY 423 CAAGTCACACGAGATTTTGCAGAAATATGATTCATCCCTCTGATGTCGAGAGAAAGTTA 482
 DB 170 GlnValaenGlnAspIleCysLysTyrGlySerIleProProAspValGlnGlnLysLeu 189
 QY 483 CGGTGAATGGCCCTACAGAGCAATTTGCTACTCCGAGAACATACACAGAAAGAAAGTTG 542
 DB ArgLeuGlnTrpProTyrGlnGlnGlnLeuLeuAryGlnLysTyrGlnLysSerPhe 209
 QY 543 AAGAACAGACCTACTCAAGAAAGCTCTGTAGAATGCTATACATTTTGCAGAACTGCTCA 602
 DB LysAenSerThrTyrSerArgSerSerValAspValLeuTyrThrPheAlaAenCysSer 229
 QY 603 GGACTGCACTGATCTTTGGCTTAATGCTTATTAAGAAGCAGATTTGAGTGGAGAAC 662
 DB 220 GlyLeuAepLeuIlePheGlyLeuAenAlaLeuLeuAryGlnAlaAspLeuGlnTrpAsn 249
 QY 663 AGTTCAATGCTCACTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 722
 DB SerSerAenAlaGlnLeuLeuLeuAspTyrCysSerSerLysGlyTyrAsnIleSerTrp 269
 QY 723 GAACCTAGCAATGAACCTTAACAGTTTCTTAAGAGGCTGATTTTCAATAGGATTCG 782
 DB 270 GlnLeuGlyAenGlnProAenSerPheLeuLysLysAlaAspIlePheIleAenGlySer 289
 QY 783 CAGTTAGAGAAAGATTATTAATTCATTAATTCATTAATTCATTAATTCATTAATTCAT 842
 DB 290 GlnLeuGlyGlyAspTyrIleGlnLeuHisLysLeuLeuAryLysSerThrPheLysAsn 309
 QY 843 GCAAAACCTATGATGCTGATGTTGGTGTGAGCTGCAAGAGAGCGCTTAAGTGTGAG 902
 DB 310 AlaLysLeuTyrGlyProAenPvalGlyGlnProAryGlnLysAlaLysLeuLeuLys 329
 QY 903 AGCTTCTGAGAGGCTGTGAGAGAACTGATTAATTCATTAATTCATTAATTCATTAATTC 962
 DB 330 SerPheLeuLysAlaGlyGlyGluValIleAepSerValIleThrPheHisIleTyrTrpLeu 349
 QY 963 AATGACGAGACTGCTACACGAGAGAAATTTTCTAAACCTGATGATTTTGACATTTTAA 1022
 DB 350 AenGlyAryGlnAlaThrAryGlnAspPheLeuAenProAspValLeuAspIlePheIle 369
 QY 1023 TCATCTGACAAAGATTTTCCAGGTGATGAGAGACACAGCGCTGCGAGAGAGGCTGG 1082
 DB 370 SerSerValGlnLysValPheGlnValValGlnSerThrAryProGlyLysLysValTrp 389
 QY 1083 TTAGAGAAACAAAGCTTGCATATATGAGCGAGCGCGCTTGTCTATCCGACACTTTGCA 1142

[illegible]

DR	MP1: 1999-494300/41.
DR	N-PsDB; AAK86671
PT	New heparanase polypeptide useful for treating autoimmune diseases, skin
PT	diseases, cardiovascular diseases and nervous system diseases including
PR	Alzheimer's disease.
XX	
XX	
PS	Claim 3; Page 29-31; 40pp; English.
CC	The present sequence represents a polypeptide with human heparanase
CC	biological activity. Antagonists and inhibitors of the protein prevent it
CC	from degrading the extracellular matrix and releasing heparan sulfate
CC	from the extracellular matrix surface. The heparanase protein or the anti-
CC	heparanase antibody are used in pharmaceutical compositions for treating
CC	warm blooded animals suffering from a disease resulting from shortage or
CC	lack of the heparanase protein, or from excessive activity or over-
CC	expression of the heparanase protein, respectively. The heparanase
CC	protein is used in treating diseases such as trauma, autoimmune disease,
CC	skin diseases, cardiovascular diseases and nervous system diseases
CC	including Alzheimer's disease resulting from shortage or lack of
CC	polypeptide. The anti-heparanase antibody is used in treating the
CC	diseases like cancer, cancer metastasis, angiogenesis and inflammation
CC	including arthritis resulting from excessive activity or over expression
CC	of heparanase protein. The anti-heparanase antibody can be used to detect
CC	the presence or absence of polypeptide and its concentration. (Updated on
CC	20-MAR-2003 to correct PA field.)
XX	
XX	
Sequence 588 AA;	
Alignment Scores:	
Pred. No.:	8.75e-279 Length: 588
Score:	2904.00 Matches: 553
Percent Similarity:	100.00% Conservative: 1
Best Local Similarity:	99.82% Mismatches: 0
Query Match:	93.11% Indels: 0
DB:	2 Gaps: 0
US-10-676-079-3 (1-1721) x AAY30124 (1-588)	
QY	30 AGCTGGCGGGGGAGACAGCAGGTGAGCCCAAGATGCTGCTCGCTCGAAGCCTGCGCTG 89
DB	35 SERTPAAGGIGYIGUGInProGlyuInuProlySmellLeuLuarSerlyProAlaLeu 54
QY	90 CCGCCGCGCGTGATGCTGCTGCTCCTGGGGCGCGTGCGGTCCCTCTCCCTGGCGCGCTG 149
DB	55 ProProPoleuMetLeuLeuLeuLeuGlyProLeuGlyProLeuSerProGlyAlaLeu 74
QY	150 CCGGACGTCGGCGCAAGCAAGAGCGTCGGAGCCTTCTTCCACCCAGGAGCGCGTG 209
DB	75 ProHapProAlaGlnAlaGlnAlaHapAlaValaHapLeuAspPhePheTrinGlnuProLeu 94
QY	210 CACCTGTAGAGCCCTCGTTCCTGTCGTCGTCACATTGAGCCCAACTGGCCAGCGACCG 269
DB	95 HisLeuValSerProSerPheLeuSerValThrileAspAlaAsmLeuAlaThrAspPro 114
QY	270 CGGTTCCCTCATCCCTCGGTTCTCCAAAGCTTCGTTACCTGGCCAGAGCGTTCCTCT 329
DB	115 ArgPheLeuIleLeuLeuGlySerProLybLeuArgThrLeuAlaArgIlyeUsSerPro 134
QY	330 GCGTACCTGAGGTTGGTGGTGACCAAGACAGACTTCTTAATTTGGATCCCAAGAGAA 389
DB	135 AlaTryLeuAArgPheGlyGlyThrlyrThrAspPheLeuIlePheAspProLybysGln 154
QY	390 TCAACCTTTGAAGAGAAAGTTACTGCGCAATCTCAAGTCAACCGAGATTTTGGCAATYT 449
DB	155 SerThrPheGlnGluArgSerlyrTrpGlnSerGlnValaAsnGlnAspIleCyblyrTy 174
QY	450 GGATCCATCCCGCCGATGTGGAGAGAAAGTTACGTTGGATTTGGATGGCCACAGAGGCAA 509
DB	175 GlySerIleProProAspValaGlnuLybLeuAArgLeuGlnuTrpProLyrgInGlnuGln 194
QY	510 TTGCTACTCGAGAACTACTACGAAAAAGTTCAAGAAACAGCACTTACTCAAGAACTCT 569


```

Db      195  LeuLeuLeuArgGluH1eTyGlnLysLysPheLysAaenSerThrTySerArgSerSer 214
Qy      570  GTAGATGGCTATACCTTTGGCAACCTGCTGAGACCTGATCTTGGCCATAAT 629
Db      215  ValAlaPheValLeuTyThrPheAlaAenCySerGlyLeuApeLeuIlePheIlyLeuAen 234
Qy      630  GCGTATTAAAGAACAGATTGTCAGTGAACAGTTCTAAAGCTCAGTTGCTCTGGAC 689
Db      235  AlaLeuLeuArgThrAlaApeLeuGlnTrpAenSerSerAenAlaGlnLeuLeuAen 254
Qy      690  TACTGCTCTTCCAGAGGGTATACATTTCTGGGAACCTAGCAATGAACCTACAGTTTC 749
Db      255  TyrCySerSerSerGlyTyAenIleSerTrpGlnLeuGlyAaenGlnProAenSerPhe 274
Qy      750  CTTAAGAAGCGCTGATATTTCATCAATGGTGGCGATTAGGAAGATTAATTCATTTG 809
Db      275  LeuLysLysAlaApeIlePheIleAenGlySerGlnLeuGlyAlaApePheIleGlnLeu 294
Qy      810  CATTAACCTTCTAAAGAAAGTCCACCTTCAAAAATGCAAACTGATAGCTCTGATTTGGT 869
Db      295  HisLysLeuLeuArgLysSerThrPheLysAenAlaLysLeuTyGlyProApeValGly 314
Qy      870  CAGCTCGAAGAAGACGGCTAAGATGCTGAAGAGCTTCTGAGAGCTGCTGAGAAAGTG 929
Db      315  GlnProArgArgLysThrAlaLysMetLeuLysSerPheLeuLysAlaGlyGlyVal 334
Qy      930  ATTGATTCAGTTACATGGCATCATACTATTGGAATGACCGGACCTGACCGGAAGAT 989
Db      335  IleApeSerValThrTrpHisIleTyTyLysAenGlyArgThrAlaThrArgGlnApe 354
Qy      990  TTTCTAAACCTGATGATTTGACATTTTATTTATTCATCTGTCGCAAAAAGTTTCCAGGTG 1049
Db      355  PheLeuAenProApeValLeuApeIlePheIleSerSerValGlnLysValPheGlnVal 374
Qy      1050  GTTGAAGACACCAAGGCTGCGCAAGAGGTCTGTTAGGAAGAAACAAGCTCTCATATGGA 1109
Db      375  ValGlnSerThrArgProGlyLysLysValTrpLeuGlyGlnThrSerSerAlaTyGly 394
Qy      1110  GGGGAGAGCGCGCTTGATCCGACACCTTTGACAGCGGCTTATGCGGTGATTAATTTG 1169
Db      395  GlyGlyAlaProLeuLeuSerApeThrPheAlaAlaGlyPheMetTrpLeuApeLysLeu 414
Qy      1170  GGCCTGTCAAGCCGGAATGGAATAGAGGTGATGAGGCAAGTATCTTTGAGACAGGA 1229
Db      415  GlyLeuSerAlaArgMetCylIleGlyValIleMetArgGlnValPhePheGlyAlaGly 434
Qy      1230  AACTACCATTTAGTGATGAAAACTTGATCTTTAAGTATTTAGCTTATCTCTTG 1289
Db      435  AenTyThrIleLeuValApeGlnAenPheApeProLeuProApeTyTrpLeuSerLeuLeu 454
Qy      1290  TTCAAGAAATTTGGTGGGCAACAGGTGTTAATGGCAAGGTGCCAAGTTCAAGAGAGG 1349
Db      455  PheLysLysLeuValGlyThrLysValIleMetAlaSerValGlnGlySerLysArgAag 474
Qy      1350  AAGCTTCGAGTATTAACCTTCATTGACACAACACATGCAATCCAAAGGTATTAAGAGAGAT 1409
Db      475  LysLeuAagValTyLysLeuHisCylThrAenThrApeAenProArgTyLysGlnGlyApe 494
Qy      1410  TTAACCTGTATGCCATTAACCTCCATTAAGTCAACCAAGTACTTGGGTTACCTTACT 1469
Db      495  LeuThrLeuTyAlaIleAenLeuHisAenValThrLysTyLysLeuArgLeuProTyPro 514
Qy      1470  TTTTCTTAAGACAGATGATTAATTAATCTTTAAGACCTTTGGAGCTCATGATTAATCT 1529
Db      515  PheSerAenLysGlnValApeLysTyLysLeuLysApeProLeuGlyProHisGlyLeuLeu 534
Qy      1530  TCCAAATCTGTCCAATCTCAATGGTCTAATCTTAAGATGGATGATGATAAAGCTTGCCA 1589
Db      535  SerLysSerValGlnLeuAenGlnLysLeuThrLysLeuMetValApeApeGlnThrLeuPro 554
Qy      1590  CCTTATATGAAAAACCTCTCGGCGCAAGAGTTCATCGGCGCTTGCGACCTTCTCATAT 1649
Db      555  ProLeuMetGlnLysProLeuArgProGlySerSerLeuGlyLeuProAlaPheSerTyx 574

```

```

Qy      1650  AGTTTTTGTGATTAAGAAATGCCAAAGTGTCTGTCATC 1691
Db      575  SerPhePheValIleArgAenAlaLysValAlaAlaCysIle 588

```

RESULT 4

AAV02345

ID AAV02345 standard; protein; 543 AA.

AC AAV02345;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

```

US-10-676-079-3 (1-1721) x AAV02345 (1-543)

```

Sequence 543 AA;

Alignment Scores:

Pred. No.: 1.2e-272

Score: 2842.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 91.12%

DB: 2

Gaps: 0

Length: 543

Matches: 543

Conservative: 0

Mismatch: 0

Indels: 0

Gaps: 0

Claim 6; Fig 1; 63pp; English.

The specification describes a polypeptide having heparanase (hp) activity. The recombinant protein is used as a modulator of heparin-binding growth factors, cellular responses to heparin-binding growth factors and cytokines, cell interaction with plasma lipoproteins, cellular susceptibility to viral, protozoal and bacterial infections or conditions such as wound healing, angiogenesis, restenosis, atherosclerosis, inflammation, neurodegenerative diseases, and viral infections. Mammalian heparanase can be used to neutralize plasma heparin, and anti-heparanase antibodies may be applied for immunodetection and diagnosis of micrometastases, autoimmune lesions, and renal failure in biopsy specimens, plasma samples, and body fluids. The present sequence represents human heparanase

Pecker I, Vlodaevsky I, Feinstein E;

WPI, 1999-30225/25.

DR N-PSDB; AAX35648.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX MPI: 2000-062144/05.
 DR N-PSDB; AA239195.
 XX Engineered cells that express recombinant heparanase, useful
 PT therapeutically, e.g. for treating angiogenesis and to screen for
 PT specific inhibitors, potential anticancer agents.
 XX Claim 3; Page 107-109; 118pp; English.
 PS
 CC The present invention describes genetically modified cells (A) containing
 CC a polynucleotide (I) that encodes a polypeptide with heparanase activity,
 CC and express recombinant heparanase (II). Heparanase cleaves heparan
 CC sulphate (HS) at specific intrachain sites, resulting in release of
 CC heparin-binding growth factors, enzymes and proteins that are sequestered
 CC by HS in basement membranes, extracellular matrix or cell surfaces. It
 CC may also be implicated in tumour angiogenesis and metastases. (II) is
 CC potentially useful in wound healing and for treating angiogenesis,
 CC restenosis, atherosclerosis, inflammation, neurodegeneration, viral
 CC infection and cystic fibrosis. It can also be used to neutralise heparin
 CC (an alternative to protamine) and to screen for specific inhibitors
 CC (potentially useful for treating cancer and metastases). Antibodies
 CC raised against (II) are used for immunodetection and diagnosis of
 CC micrometastases, autoimmune lesions and kidney failure. (A) provide (II)
 CC in large quantities, in a form that is homogeneously processed and
 CC activated/neutralised by a dedicated protease. The present sequence
 CC represents human heparanase
 XX
 XX Sequence 543 AA:
 SQ
 Alignment Scores:
 Pred. No.: 1 2e-272 Length: 543
 Score: 2842.00 Matches: 543
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 91.12% Indels: 0
 DB: 3 Gaps: 0
 US-10-676-079-3 (1-1721) x AA57590 (1-543)
 QY 63 ATGCTGCTGCGCTGAGAGCTGCTGCGCGCGCTGATGCTGCTGCTGCGGCGG 122
 DB 1 MetLeuLeuArgSerIysProAlaLeuProProLeuMetLeuLeuLeuGlyPro 20
 QY 123 CTGGGCTCCCTCTCCCTGCGCGCTGCGCGCGCTGCGGAGCAAGACGCTGCGAC 182
 DB 21 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaAspValValaAsp 40
 QY 183 CTGGACTTCTTACCCGAGAGCGCGCTGCACTGCTGAGCGCGCTGCTGCTGCTGCTG 242
 DB 41 LeuAspPhePheThrGlnGluProLeuHisLeuValSerProSerPheLeuSerValThr 60
 QY 243 ATTGACGCAACCTGCGCAGAGCGCGGCTTCTGATCTCTCTGAGTTCTCCAAAGCTT 302
 DB 61 IleAspAlaAsnLeuAlaThrAspProArgPheLeuIleLeuLeuGlySerProIysLeu 80
 QY 303 CGTACCTGGCCAGAGGCTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 362
 DB 81 ArgThrLeuAlaArgGlyLeuSerProAlaTyrLeuArgPheGlyGlyThrIysThrAsp 100
 QY 363 TTCTCTAATTTGATCCCAAGAGAACTTGAAGAGAGAAAGTTACTGCAATCT 422
 DB 101 PheLeuIlePheAspProIysIysGlySerThrPheLeuGlnArgSerTyrTrpGlnSer 120
 QY 423 CAAGTCAACGAGATATTTGCAATATGATCCATCTCTGATGCTGAGAGAGAAAGTTA 482
 DB 121 GlnValAsnGlnAspIleGlySerTyrGlySerIleProProAspValGlnGlnIysLeu 140
 QY 483 CGGTGGATGCGCCCTTCAACGAGCAATGCTATCCGAGACATCAACGAAAGAGTTTC 542
 DB 141 ArgLeuGlnIuprProIyGlnGlnGlnLeuLeuArgGlnIuprIyGlnIysIysPhe 160
 QY 543 AAGAAACGACCTACTCAAGAGCTCTGATGATGTGTATACACTTTTGAAAGCTGTCA 602

DB 161 LysAsnSerThrTyrSerArgSerSerValAspValLeuTyrThrPheAlaAsnGlySer 180
 QY 603 CGACTGCACTTGATCTTTGGCTTAAGCCCTTATTAAGACAGAGATTTGACATGGAGAC 662
 DB 181 GlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGlnThrAsn 200
 QY 663 AGTCTTAATGCTCAGTTGCTCTCTGAGTACTGCTCTTCCAAAGGAGTAAATCATTTCTTGG 722
 DB 201 SerSerAsnAlaGlnLeuLeuLeuLeuAspTyrCysSerSerIysGlyTyrAsnIleSerThr 220
 QY 723 GAAGTGGCAATGAACTTAACAGTTTCTTAAAGAGCTGATATTTTCATCATAGGCTG 782
 DB 221 GlnLeuGlyAsnGlnProAsnSerPheLeuIysIysAlaAspIlePheIleAsnGlySer 240
 QY 783 CAGTTAGAGAAAGATTAATTCATTTGATTAAGTAACTTCGTAAGAGTCCAGCTGAAAGAT 842
 DB 241 GlnLeuGlyGlnAspTyrIleGlnLeuHisIysLeuLeuArgIysSerThrPheIysAsn 260
 QY 843 GCAAAACTCTATGCTGCTGATGTTGCTGAGCGCTGAGAAAGAGCGCTTAAGATGCTGAG 902
 DB 261 AlaIysLeuTyrGlyProAspValGlyGlnProArgGlySerThrAlaIysMetLeuIys 280
 QY 903 AGCTTCTGAAAGCTGCTGAGAGAGTGAATGATTAATCACTTACATGAGCATCACTACTATTG 962
 DB 281 SerPheLeuIysAlaGlyIyGlyValIleAspSerValThrTrpHisIleTyrTyrLeu 300
 QY 963 AATGACGCACTGCTCAACAGGAGAAATTTTCTAAACCTGATGATTTGACATTTTATTT 1022
 DB 301 AsnGlyArgThrAlaThrArgGlnAspPheLeuAsnProAspValLeuAspIlePheIle 320
 QY 1023 TCATCTGCAAAAAGTTTTCAGAGTGTGAGAGACACAGCGCTGCAAGAGAGCTG 1082
 DB 321 SerSerValGlnIysValPheGlnValAlaGlnSerThrAspProGlyIyIysValIupr 340
 QY 1083 TTAGAGAAACAAAGCTCTGCAATATGAGAGCGAGCGCTTGTCTATCCGACACTTTGCA 1142
 DB 341 LeuGlyGlnThrSerSerValTyrGlyGlyAlaProLeuLeuSerAspThrPheAla 360
 QY 1143 GCTGCTTATATGCTGCTGATTAATTTGGCGCTGTACGCCCAATGGAATTCAGAGTGTG 1202
 DB 361 AlaGlyPheMetThrLeuAspIyysLeuGlyLeuSerAlaAspGmeGlyIleGlnValVal 380
 QY 1203 ATGAGGCAAGATTTCTTTGAGAGAGAACTACCTTTAGTGAAGAACTTCGATCT 1262
 DB 381 MetArgIleValPhePheGlyAlaGlyAsnTyrHisLeuValAspGlnAspPheAspPro 400
 QY 1263 TTACTGATTAATTTGCTATCTCTCTGTTCAAGAAATTTGTTGGGACCAAGGTTGTAATG 1322
 DB 401 LeuProAspTyrTrpLeuSerSerLeuLeuPheIysIysValGlyThrIysValLeuMet 420
 QY 1323 GCAAGCGTGAAGGTTCAAGAGAGAAAGCTTGAAGTATCTTCAATTTGCAACAAACT 1382
 DB 421 AlaSerValGlnGlySerIysArgIyysLeuArgValIyTyrLeuHisCysThrAsnThr 440
 QY 1383 GACATTCAGAGTAAAGAGAGAGATTAATCTGTAATGSCATTAACCTCATAGAGTC 1442
 DB 441 AspAsnProArgTyrIyysGlnIyysAspLeuThrIyysAlaIleAsnLeuHisAsnVal 460
 QY 1443 ACCAAGTACTTGGCTTACCTTATCTTTTCTTAAAGAGCAAGTGAATATCTTCTTA 1502
 DB 461 ThrIyysTyrLeuArgLeuProIyysProPheSerAsnIyysGlnValAspIyysTyrLeu 480
 QY 1503 AGACCTTTGGGACCTTCATGATTAATCTTTCATTAATCTGTCATTAATGTTAACTCTTA 1562
 DB 481 ArgProLeuGlyProHisIyysLeuLeuSerValGlnLeuAsnGlyLeuThrIyys 500
 QY 1563 AAGATGCTGAGAGATCAACCTTGGCACCTTAAAGGAAACCTTCGCGGACAGAGAGT 1622
 DB 501 IyysMetValAspAspGlnThrIyysProProLeuMetGlnIyysProLeuArgProIyys 520
 QY 1623 TCATGCGCTTCCAGAGCTTCTCATATAGTTTCTTTGTGATTAAGAAATGCAAGTGTCT 1682

Db 521 SerLeuGlyLeuProAlaPheSerTyrSerPhePheValIleArgAlaLysValAla 540
 QY 1683 GCTTGATC 1691
 Db 541 AlaCysIle 543

RESULT 6
 AAB08849
 ID AAB08849 standard; protein, 543 AA.
 AC AAB08849;
 XX
 DT 15-JAN-2001 (first entry)
 DE Amino acid sequence of a human heparanase polypeptide.
 XX
 KM Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
 KM heparin-binding growth factor; cytokine; neurodegenerative plaque;
 KM wound healing; infection; burn; angiogenesis; restenosis;
 KM atherosclerosis; inflammation; neurodegenerative disease;
 KM Gerstmann-Strausler Syndrome; Creutzfeldt-Jakob disease.
 XX
 OS Homo sapiens.
 XX
 PN WC020052178-A1.
 XX
 PD 08-SEP-2000.
 XX
 PF 14-FEB-2000; 2000WO-US003542.
 XX
 PR 01-MAR-1999; 99US-00258892.
 XX
 PA (INST-) INSIGHT STRATEGY & MARKETING LTD.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 PA (FRIE/) FRIEDMAN M M.
 XX
 PI Pecker I, Vlodaevsky I, Feinstein E;
 XX
 DR WPI, 2000-579289/54.
 DR N-PSDB; AAA75051.
 XX
 PT New polynucleotides encoding a polypeptide having heparanase activity,
 PT useful in wound healing and in gene therapy, particularly in treating
 PT tumor, inflammation, autoimmunity, neurodegenerative diseases.
 XX
 PS Claim 22; F4g 1; 152pp; English.
 XX
 CC The present sequence represents a human protein with heparanase catalytic
 CC activity. The heparanase (hpa) polynucleotide is useful in gene therapy,
 CC particularly in treating tumour, inflammation or autoimmunity.
 CC Particularly, the polynucleotide is useful in modulating the
 CC bioavailability of heparin-binding growth factors, cellular responses to
 CC heparin-binding growth factors (e.g. bFGF) and cytokines (e.g.
 CC interleukin (IL)-8), cell interaction with plasma lipoproteins, cellular
 CC susceptibility to certain viral and some bacterial and protozoa
 CC infections, or disintegration of neurodegenerative plaques. The
 CC polynucleotide is also useful in wound healing (e.g. thermal, chemical or
 CC radiation burns), and in the treatment of angiogenesis, restenosis,
 CC atherosclerosis, inflammation, neurodegenerative diseases (Gerstmann-
 CC Strausler Syndrome or Creutzfeldt-Jakob disease), and some viral,
 CC bacterial or protozoa infections
 XX
 SO Sequence 543 AA;

Alignment Scores:
 Pred. No.: 1.2e-272 Length: 543
 Score: 2842.00 Matches: 543
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 91.12% Indels: 0
 DB: 3 Gaps: 0

US-10-676-079-3 (1-1721) X AAB08849 (1-543)

QY 63 ATGCTGCGGCTCGAAGCCCTGCGCGCGCGCTGATGCTGCTCTGGGGCCG 122
 Db 1 MetLeuLeuArgSerLysProAlaLeuProProLeuMetLeuLeuLeuGlyPro 20
 QY 123 CTGGATCCCTCTCCCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182
 Db 21 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaGlnAspValAlaAsp 40
 QY 183 CTGGAATCTTCAACCGAGAGCGCGCTGCACTGGTAGCCCTGTTCTGTCCTGAC 242
 Db 41 LeuAspPhePheThrGlnGlnProLeuHisLeuValSerProSerPheLeuSerValThr 60
 QY 243 ATTAGCGCCCACTGGCCACGAGACCGCGGTTCTCTAATCTCTGAGTTTCCAAAGCTT 302
 Db 61 IleAspAlaAsnLeuAlaThrAspProArgPheLeuIleLeuLeuGlySerProLysLeu 80
 QY 303 CGTACCTGGCGCAGAGGCTGTGCTCCGCTGCTGAGGTTGGTGGCACCAAGACAGAC 362
 Db 81 ArgThrLeuAlaArgGlyLeuSerProAlaTyrLeuArgPheGlyGlyThrLysThrAsp 100
 QY 363 TTCCTAATTTTCGATCCCAAGAGATCAACCTTGAAGAGAGATTACTGGCAATCT 422
 Db 101 PheLeuIlePheAspProLysLysGlnSerThrPheGlnGlnArgSerTyrTrpGlnSer 120
 QY 423 CAAGTCAACCGAGATTTTGCAGAAATATGATCCATCCCTCTGATGCGAGGAAGTTA 482
 Db 121 GlnValAsnGlnAspIleCysLysTyrGlySerIleProProAspValGlnGlnLysLeu 140
 QY 483 CGGTGGAAATGGCCCTACCGAGAGCAATGCTCTCCGAGAACATCCAGAAAGTTTC 542
 Db 141 ArgLeuGlnTrpProLysGlnGlnGlnLeuLeuAspGlnHisTyrGlnLysLysPhe 160
 QY 543 AAGAACAGCACTACTCAAGAGAGCTGTAGATGCTATACACTTTGGCAACCTGCTCA 602
 Db 161 LysAsnSerThrTyrSerArgSerSerValAspValLeuTyrThrPheAlaAsnGlySer 180
 QY 603 GGACTGGAATGATCTTTGGCTTAATGCGTTTAAAGAACAGAGATTGGAGTGAAC 662
 Db 181 GlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrLysAspLeuGlnTrpAsn 200
 QY 663 AGTTCTAATCTCAGTTGCTCCGAGCTAAGCTAAGGCTGATTAATTTCTGG 722
 Db 201 SerSerAsnAlaGlnLeuLeuLeuAspTyrCysSerSerLysLysLysTrp 220
 QY 723 GAACTAGGCAATGAACCTAACAGTTTCTTAAGAGGCTGATTTTCAATGAGGTCG 782
 Db 221 GlnLeuGlyAsnGlnProAsnSerPheLeuLysLysAlaAspIlePheIleAsnGlySer 240
 QY 783 CAGTTAGAGAGATTTTATTTCAATTGCAATTAACCTTAAAGAGTCCACTTCAAAAT 842
 Db 241 GlnLeuGlyGlnAspTyrIleGlnLeuHisLysLeuLeuAspGlySerThrPheLysAsn 260
 QY 843 GCAAACTCATGAGTCTGATGTTGGTCAAGCTCGAAGAAAGCGGCTAAGATGCTGAAG 902
 Db 261 AlaLysLeuTyrGlyProAspValGlyGlnProArgLysThrAlaLysPheLeuLys 280
 QY 903 AGCTTCTGAGGCTGAGAGAGATGATTGATTCAGTTACATGAGCATCACTAATTTG 962
 Db 281 SerPheLeuLysAlaGlyGlyGlnValIleAspSerValThrTrpHisLysTyrTrpLeu 300
 QY 963 AATGACGAGCTCTACCGAGGAAGATTTTCTTAACCTCGATGATTTGGACATTTTAT 1022
 Db 301 AsnGlyArgThrLysThrArgGlnAspPheLeuAsnProAspValLeuAspIlePheIle 320
 QY 1023 TCATCTGTGGAAGAAATTTTCCAGGCTGTGAGAGACCAAGGCTGGCAAGAGGCTGG 1082
 Db 321 SerSerValGlnLysValPheGlnValValGlnSerThrArgProLysLysValTrp 340
 QY 1083 TTAGAGAAACAAGCTCTGATATGAGAGCGAGCGCTTGTCTATCCGACACTTTGCA 1142
 Db 341 LeuGlyGlnThrSerSerLysLysTyrGlyGlyAlaProLeuLeuSerAspThrPheAla 360


```

Db      181 GlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGlnTrpAsn 200
QY      663 AGTTCTATGCTCAGTTCCTCTGACTACTGCTCTTCCAAAGGGGATTAACATTTCTTGG 722
Db      201 SerSerSerAsnAlaGlnLeuLeuLeuLeuAspTyrCysSerSerTyrSgLYrAsnIleSerTyr 220
QY      723 GAACCTAGGCAATGAACCTTAACAGTTTCCTTAAGAGCGTGATATTTTCATCAATGCGTTCG 782
Db      221 GluLeuGlyAsnGlnProAsnSerPheLeuSylAsAlaPheIleAsnGlySer 240
QY      783 CAGTTAGAGAGAGATTTATTCATTCATTAACCTTCAAGAAAGTCCACCTTCAAAAAT 842
Db      241 GlnLeuGlyGluAspTyrIleGlnLeuHisIlyLeuLeuArgLysSerThrPheLysAsn 260
QY      843 GCAAAACTATATGCTCTGATGTGTGTCAGGCTCGAAGAAAGACGCGTAAAGATGCTGAG 902
Db      261 AlaLysLeuTyrGlyProAspValGlyGlnProArgTrgIlyThrAlaLysMetLeuLys 280
QY      903 AGCTTCTGAAGCGTGTGAGAAAGTGAATTGATTCAGTTACATGGCATCACTACTATTGG 962
Db      281 SerPheLeuSylAsnGlyGlyGlnValIleAspSerValThrTrpHisIlyTyrLeu 300
QY      963 AATGAGAGGACGTGTACAGGAGAGATTTTCTAAACCTGATGATGATGACATTTTATT 1022
Db      301 AsnGlyArgThrAlaThrArgGluAspPheLeuAsnProAspValLeuAspIlePheIle 320
QY      1023 TCATCTGTGCAAAAAGTTTTCAGAGTGTGAGAGCAACGAGCTGCGCAAGAGTCTGG 1082
Db      321 SerSerValGlnLysValPheGlnValValGlnSerThrArgProGlyLysValTyr 340
QY      1083 TTAGAGAGAAACAGCTCTGCATATGAGAGCGAGCGCTTGTCTATCCGACCTTTGCA 1142
Db      341 LeuGlyGluThrSerSerAlaTyrGlyGlyValaProLeuLeuSerAspThrPheAla 360
QY      1143 GCTGCTTTATGTTGCTGATTAATTTGGGCTGTACGCGCAATGGGAATGAAAGTGTG 1202
Db      361 AlaGlyPheMetTrpLeuAspLysLeuGlyLeuSerAlaArgMetGlyIleGlnValVal 380
QY      1203 ATGAGCGAAGTATTCTTTGAGACAGAAACCTACATTAGTGAATGAAAATTGATCTT 1262
Db      381 MetArgGlnValPhePheGlyAlaGlyAsnTyrHisLeuValaAspGluAsnPheAspPro 400
QY      1263 TTACCTGATTATTTGGCTATCTCTCTGTTCAGAAATTTGGTGGCACCAAGCTTTATG 1322
Db      401 LeuProAspTyrTrpLeuSerLeuLeuPheLysLeuValGlyTyrTrpValLeuMet 420
QY      1323 GCAAGCGTGAAGGTTCAAGAGAGAAAGCTTGAGTATACCTTCACTTGACAAAACCT 1382
Db      421 AlaSerValGlnGlySerLysArgArgLysLeuArgValTyrLeuHisCysThrAsnThr 440
QY      1383 GACAACTCAAGGTATTAAGAAGAGATTTAACTCTGTATGCGCATTAACCTCCATTAACGTC 1442
Db      441 AspAsnProArgTyrLysGlnGlyAspLeuThrLeuTyrAlaIleAsnLeuHisAsnVal 460
QY      1443 ACCAACTACTTGGCGGTACCTATCTCTTTTCTAACAGCAAGTGGATTAATACCTTCTA 1502
Db      461 ThrLysTyrLeuArgLeuProLysProPheSerAsnLysGlnValaAspLysTyrLeuLeu 480
QY      1503 AGACCTTTGGGACCTCATGATTACTTTCCAAATCTGCGCAACTCATGAGCTTAACCTTA 1562
Db      481 ArgProLeuGlyProHisGlyLeuLeuSerLysSerValGlnLeuAsnGlyLeuThrLeu 500
QY      1563 AAGATGTGATGATCAACCTTGCCACCTTTAAATGAAAAAAGCTCTCGGCGCAGAGAGT 1622
Db      501 LysMetValaAspArgGlnThrLeuProProLeuMetGlyLysProLeuArgProGlySer 520
QY      1623 TCATCTGGGCTTGGCAGCTTCTCATATAGTTTTTTTGTGATAGAAATGCCAAGTTGCT 1682
Db      521 SerLeuGlyLeuProAlaPheSerTyrSerPhePheValIleArgAsnAlaLysValaIla 540
QY      1683 GCTTGATC 1691
Db      541 AlaCysIle 543

```

```

RESULT 8
AA97635
ID AA97635 standard; protein; 543 AA.
AC
XX
XX
AC AA97635;
XX
XX
DT 20-APR-2001 (first entry)
XX
DE Human heparanase protein sequence.
XX
KW Heparanase; hnp1; wound healing; angiogenesis; restenosis; Scarpe;
KW atherosclerosis; inflammation; pulmonary disease; Alzheimer's disease;
KW neurodegenerative disease; Creutzfeldt-Jakob disease; viral infection;
KW gene therapy; human.
XX
OS Homo sapiens.
XX
PN WO200100643-A2.
XX
PD 04-JAN-2001.
XX
PF 19-JUN-2000; 2000MO-IL000358.
XX
PR 25-JUN-1999; 99US-0140801P.
XX
PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX
PI Pecker I, Michal I, Itzhaki H;
XX
DR WPI, 2001-137930/14.
XX
PT New polynucleotides and polypeptides that are distantly homologous to
PT heparanase, useful in wound healing, as well as in gene therapy protocols
PT for angiogenesis, restenosis, atherosclerosis, or inflammation.
XX
PS Disclosure; Page 64-65; 67pp; English.
XX
CC This sequence represents a heparanase of the invention. The heparanase
CC DNA and protein sequences are useful in wound healing, angiogenesis,
CC restenosis, atherosclerosis, inflammation, pulmonary diseases,
CC neurodegenerative diseases (such as Scarpe, Alzheimer's disease, and
CC Creutzfeldt-Jakob disease) or viral infections. The heparanase coding
CC sequence is particularly useful in gene therapy
XX
SQ Sequence 543 AA:
Alignment Scores:
Pred. No.: 1.2e-272 Length: 543
Score: 2842.00 Matches: 543
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 91.12% Indels: 0
DB: 4 Gaps: 0
US-10-676-079-3 (1-1721) x AA97635 (1-543)
QY 63 ATGCTGCTGAGCTCAAGACCTGCGCTGCGCGCGCTGATGCTGCTGCTGCGGCGG 122
Db 1 MetLeuLeuArgSerLysProAlaLeuProProProLeuMetLeuLeuLeuGlyPro 20
QY 123 CTGGGTCCTCTCTCCCTGCGCGCTGCGCGCGCTGCGCAAGCACAGGACCTGCGAGC 182
Db 21 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaAspValaAsp 40
QY 183 CTGGACTTCTTCAACCAAGAGCGCGCTGCGACTGAGTGAAGCCCTCTGTCCTCCGAC 242
Db 41 LeuAspPhePheThrGlnGluProLeuHisLeuValaSerProSerPheLeuSerValaThr 60
QY 243 ATGAGCGCAACCTGCGCACGAGACCGCGGTTCTCATCTCTGAGTTCTCCAAAGTTT 302
Db 61 IleAspAlaAsnLeuAlaThrAspProArgPheLeuIleLeuLeuGlySerProLysLeu 80

```


QY	303	CGTACCTTTGGCAAGAGCTGTCTCCCTGGCAACTGAGAGTTTGGTGGCAACAAAGACAG	362
Db	81	ArgThrIleuAlaAGGlyLeuSerProAlaTyrIleuArgPheGlyGlyThrLysThrAsp	100
QY	363	TTCTTAATTTTCGATCCCAAGAAAGAAATCAACTTTTAAAGAGAGAAAGTTACTGGCAATCT	422
Db	101	PheIleuIlePheAspProLysGlyIleSerThrPheGluGluAspGertyrTrpGlnSer	120
QY	423	CAAGTCACACGAGATATTTCGAAATATGATCCATCCCTCCTGATGTGGAGAGAAATTA	482
Db	121	GlnValAsnGlnAspIleCysLysTyrGlySerIleProProAspValGluGluLysLeu	140
QY	483	CGGTTGGAAATGGCCCTACACAGAGCAATTGTACTCCGAAACACTACACAGAAAAGTTC	542
Db	141	ArgIleuIleuTrpProTyrGlnGluGlnLeuLeuLeuArgGlnHisTyrGlnLysLysPhe	160
QY	543	AAGAACAGCACTCACTCAAGAAAGCTCTGTAGATGGCTATACACTTTTGGCAATGGCTCA	602
Db	161	LysAsnSerThrTyrSerArgSerSerValAspValIleuTyrTrnPheAlaSerCysSer	180
QY	603	GGACTGGACTTGATCTTTGGCTTAAATGGCTTATTAAGAACAGCAGATTTGGACGTGAAAC	662
Db	181	GlyLeuAspIleuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGlnTrpAsn	200
QY	663	AGTTCTTAATGTCAAGTTGCTCTCTGGACTACACTGCTCTTCCAGGGGGTATACATTTCTGG	722
Db	201	SerSerAsnAlaGlnLeuLeuLeuAspTyrCysSerSerLysGlyTyrAsnIleSerTrp	220
QY	723	GAACTAGGCAATGAACCTTAACAGTTTCCTTAAAGAGGTGATATTTTCATCAATGGGTCG	782
Db	221	GlnLeuGlyAsnGlnTrpAsnSerPheLeuLysLysAlaAspIlePheIleAsnGlySer	240
QY	783	CAGTTAGAGAGATATTATTCATTTGATTAACCTTCAAGAAAGTCCACCTTCAAAAT	842
Db	241	GlnLeuGlyGluAspTyrIleGlnLeuHisLysLeuLeuAspArgLysSerThrPheLysAsn	260
QY	843	GCAAAACTCTATGGTCTCGATGTTGGTCAGCCTCGAAGAAAGACGGCTAGATGCTGAAG	902
Db	261	AlaLysLeuTyrGlyProAspValGlyGlnProArgLysThrAlaLysMetLeuLys	280
QY	903	AGCTTCGGAAGGGCGGTGGAGAAAGTATGATCAGTTACATAGGCACTACTATTTTG	962
Db	281	SerPheLeuLysAlaGlyGlyGlnValIleAspSerValTrnTrpHisTyrTyrLeu	300
QY	963	AATGAGCGAGATGCTACACGAGAAAGATTTTAAACCTGTATGATTGACATTTTATTT	1022
Db	301	AsnGlyArgThrAlaThrArgLysAspPheLeuAsnProAspValIleAspIlePheIle	320
QY	1023	TCATCTGTGCAAAAAGTTTCCAGGTGGTTAGAGACACACAGCGCTGGCAGAGAGTCTGG	1082
Db	321	SerSerValGlnLysValPheGlnValAlaGluSerThrArgProGlyLysValAlaTrp	340
QY	1083	TTAGAGAAACAAAGCTCTGCATATGAGAGCGGACGCCCTTGATCCGACACTTGGCA	1142
Db	341	LeuGlyGlnTrnSerSerAlaTyrGlyGlyAlaProLeuLeuSerAspThrPheAla	360
QY	1143	GCTGGCTTATATGCGTGGATTAATATGGCGCTGACGCCGAATGGCAATGAATGGTG	1202
Db	361	AlaGlyPheMetTrpLeuAspLysLeuGlyLeuSerAlaAspMetGlyIleGluValAla	380
QY	1203	ATGAGGCAAGTATCTTTGGAGACAGAAACTACCATTTAGTGATGAAGAACTTGATCTT	1262
Db	381	MetArgGlnValPhePheGlyAlaGlyLysTrnHisLeuValAspGluAsnPheAspPro	400
QY	1263	TTACTCGATTATTTGGCTATCTCTTCTGTTCAGAAATTTGGTGGGACCAAGTGTTATG	1322
Db	401	LeuProAspTyrTrpLeuSerLeuLeuPheLysLeuValGlyThrLysValLeuMet	420
QY	1323	GCAAGCGTGCAGAGTTCAAGAGAGAAAGCTTCAGATATACCTTCATTTGACAAACACT	1382
Db	421	AlaSerValGlnGlySerLysArgArgLysLeuAspValTyrLeuHisCysThrAsnThr	440
QY	1383	GACAATCCAGGTATTAAGAAAGAGATTTAACTCTGATGCCATTAACCTTCATTAACGTC	1442

Db	441	AspansProArgIyrLysGIuGLIAspLeuThrlLeuYrAlaIleAsnIleuHisAsnVal	1460
Qy	1443	ACCAAGTACTGTCGGGTAAACCTATCCTTTTCTTACAACAAGATGGATTAATACCTTCTA	1502
Db	461	ThrlsTyrlLeuArgLeuProTyrProPheSerAsnLysGIuValAspLysTyrlLeuLeu	480
Qy	1503	AGACCTTTGGGAGCTCATGATTACTTTCCAAAATCTGTCCAACTCAATGCTTAACCTTA	1562
Db	481	ArgProLeuGIyProH ³ SGIyLeuLeuSerLysSerValGIuLeuAsnGIyLeuThrlLeu	500
Qy	1563	AAGATGGTGATGATCAAACTTGCGACTTAAATGGAAAAACCTCTCCGGCCAGGAAGT	1622
Db	501	LysMetValAspAspGIuThrlLeuProLeuMetGIuLysProLeuAsnProGIySer	520
Qy	1623	TCACCTGGGCTTCCGAGCTTCTTCATATAGTTTTTTTGTGATAGAAATGCCAAAGTTGCT	1682
Db	521	SerLeuGIyLeuProAlaPheSerTyrSerPhePheValIleArgAsnAlaItyValAla	540
Qy	1683	GCTTGATC 1691	
Db	541	AlaCysIle 543	
RESULT 9			
AB07813	ID	AB07813 standard; protein; 543 AA.	
XX	AC	AB07813;	
XX	DT	03-JUL-2002 (first entry)	
XX	DE	Human heparanase sequence.	
XX	KX	Heparanase; catalytic; cytosolic; antiviral; antibacterial; enzyme;	
XX	KW	anti-protoczoan; neuoprotective; heparin; human.	
OS		Homo sapiens.	
XX	Key	Location/Qualifiers	
XX	Peptide	1..35	
XX	FT	/note= "signal peptide"	
XX	FT	36..543	
XX	FT	/note= "mature protein"	
XX	PN	US2002034810-A1.	
XX	PD	21-MAR-2002.	
XX	PF	16-AUG-2001; 2001US-00930218.	
XX	PR	20-SEP-2000; 2000US-00666390.	
XX	PA	(INSI-) INSIGHT STRATEGY & MARKETING LTD.	
XX	PI	Goldsmidt O, Pecker I, Vlodayevy I, Michal I, Zeharia B;	
XX	DR	WPI; 2002-338926/37.	
XX	PT	Nucleic acid encoding avian and reptile heparanase polypeptide is useful	
XX	PT	to treat various heparin-related disorders and the signal peptide is	
XX	PT	useful in production of membrane-targeted or secreted recombinant	
XX	PT	proteins.	
XX	PS	Disclosure; Fig 1a; 39pp; English.	
XX	CC	The invention relates to an isolated avian and reptile nucleic acid,	
XX	CC	encoding a polypeptide with heparanase catalytic activity. The signal	
XX	CC	peptide of the nucleic acid can be used to express membrane-associated or	
XX	CC	secreted proteins in heterologous expression systems. The encoded	
XX	CC	polypeptides can be used to prevent tumor angiogenesis, metastasis and	
XX	CC	invasion, and to intervene with pathologies associated with impaired	
XX	CC	heparin-binding growth factors, cellular responses to heparin-binding	
XX	CC	growth factors and cytokines, cell interaction with plasma lipoproteins,	

CC cellular susceptibility to viral, protozoa and bacterial infections or
CC disintegration of neurodegenerative plaques. The present sequence
CC represents a human heparanase protein sequence used in similarity studies
XX
SQ Sequence 543 AA:

Alignment Scores:

Pred. No.:	1-26-272	Length:	543
Score:	2842.00	Matches:	543
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	91.12%	Indels:	0
DB:	5	Gaps:	0

US-10-676-079-3 (1-1721) x ABB07813 (1-543)

```

QY 63 ATGCTGCTGCGCTGGAAGCTGCGCTGCGCGCGCTGATGCTGCTGCTGCGGCGG 122
Db 1 MetleuLeuArgSerLysProAlaLeuProProProleuMetleuLeuGlyPro 20
QY 123 CTGGGTCCTCTCTCCCTGCGCGCTGCGCGCTGCGCAAGACAGACAGTGTGAC 182
Db 21 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaAspValAla 40
QY 183 CTGGAATCTTCCACGAGAGCGCGCTGACCTGTGAGCCCTGCTCTGCTGCTACC 242
Db 41 LeuAspPhePheThrGlnGlnProLeuHleuValSerProSerPheLeuSerValThr 60
QY 243 ATTGACGCCAAGCTGCGCAGCAGCAGCCGGTTCCTCATCCTCGGGTTCCTCAAGCTT 302
Db 61 IleAspAlaAsnLeuAlaThrAspProAlaPheLeuLeuLeuGlySerProLysLeu 80
QY 303 CGTACCTTGGCGCAGAGGCTGTCTCTGCGCTGAGCTTGAGTTGTGGACCAAGACAGAC 362
Db 81 ArgThrLeuAlaArgGlyLeuSerProAlaIleuArgPheGlyGlyThrIleThrAsp 100
QY 363 TTCTTATTTTCCATCCCAAGAAAGAAATCAACTTTGAAGAGAAATTTACTGCAATCT 422
Db 101 PheLeuIlePheAspProLysGlySerThrPheGlnGlnAlaGlySerTyrtPrGlnSer 120
QY 423 CAAGTCAACAGATATTTGCAATATGATTCATCCCTCTGATGTGAGAGAAAGTTA 482
Db 121 GlnValAsnGlnAspIleCybLysTyrcLysIleProProAspValGlnGlnLysLeu 140
QY 483 CGGTGGAATGCGCTTACCAAGAGCAATGCTACTCCGAGAAACCTACCAAGAAAGTTTC 542
Db 141 ArgLeuGlnIleProLysGlnGlnGlnLeuLeuArgGlnIleSerTyrcLysLysPhe 160
QY 543 AAGAAGCAGCACTTCTGAAGAGCTGTGATGATGCTATACATTTTGCAAACTGCTCA 602
Db 161 LysAsnSerThrTyrsSerArgSerSerValAspValLeuTyrtPrPheAlaAsnCybSer 180
QY 603 GGAAGTGAATGATCTTGGCGCTTAATAGCGTTATTAAGAACAGACAGATTGCAAGTGAAC 662
Db 181 GlnLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGlnIlePheAsn 200
QY 663 AGTTCTAATGCTGATGCTCTCTGACTACTGCTCTTCCAAAGGGATTAACATTTCTTGG 722
Db 201 SerSerAsnAlaGlnLeuLeuAspTyrcySerSerLysGlyTyrcAsnIleSerTyrtPr 220
QY 723 GAATAGGCAATGAACCTTACAGTTTCTTAAGAAAGCTATATTTTCATCAATAGGCTCG 782
Db 221 GlnLeuGlyAsnGlnProAsnSerPheLeuLysAlaAspIlePheIleAsnGlySer 240
QY 783 CAGTTAGAGAAAGATATATTCATATTCATTAACCTTAAAGAAAGTCCACCTTCAAAAT 842
Db 241 GlnLeuGlyGlnAspTyrtIleGlnLeuHleuLysLeuLeuArgLysSerThrPheLysAsn 260
QY 843 GCAAAATCTATGCTCTGATGTTGGTCAAGCTTGAAGAAAGACGGCTTAAGATGCTGAG 902
Db 261 AlaIleLeuTyrcLysProAspValGlnIleProArgArgLysThrAlaLysMetLeuLys 280
QY 903 AGCTTCTGAAGGCTGATGGAAGAAAGTATGATTCAATTACATGAGCACTACTATTG 962

```

```

Db 281 SerPheLeuLysAlaGlyGlyLysValIleAspSerValThrThrIleSerTyrcLysLeu 300
QY 963 AATGACGCGACTGCTACCAAGGAAAGATTTTCTAAACCTGATGATTTGACATTTTATTT 1022
Db 301 AsnGlyArgThrAlaThrArgGlnAspPheLeuAsnProAspValLeuAspIlePheIle 320
QY 1023 TCATCTGTGCAAAAGATTTTCCAGGTGTTGAGAGCAGCAGGCTGGCAAGAAAGTCTGG 1082
Db 321 SerSerValGlnLysValPheGlnValValGlnSerThrArgProGlyLysLysValTrp 340
QY 1083 TTAGAGAAACAAAGCTCTGATATGAGAGCGAGCGCCCTTGCTATCCGACACTTTGCA 1142
Db 341 LeuGlyGlnIleSerSerAlaTyrcLysGlyGlyAlaProLeuLeuSerAspThrPheAla 360
QY 1143 GCTGGCTTATATGCTGATATAATTGGGCTGTCAAGCCGAATGGGAAATAGAGTGTG 1202
Db 361 AlaGlyPheMetThrLeuAspLysLysGlyLeuSerAlaArgMetCylIleGlnValAla 380
QY 1203 ATGAGCGAAGTATCTTTTGAGAGGAAACCTACCATTTAGTGAGTAAACCTTCGATCCT 1262
Db 381 MetArgGlnValPhePheGlyAlaGlyAsnTyrcLysValLeuValAspGlnAsnProAsp 400
QY 1263 TTACCTGATTAATGCGTATCTCTGTTCAAGAAATGTGGCGCAAGGTTAAAG 1322
Db 401 LeuProAspTyrtPrLeuSerLeuLeuPheLysLysValGlyThrLysValLeuMet 420
QY 1323 GCAAGCGTCAAGGTTCAAAAGAGAAAGAGCTTCAGTATACCTTCATTGCAACAAACAT 1382
Db 421 AlaSerValGlnIleSerTyrcLysArgGlyLysLeuArgValTyrcLysIleSerThrAsnThr 440
QY 1383 GACAAATCAAGTATTAAGAAAGAAAGTAACTGTGATGCCATAAACCTCCATAACGTC 1442
Db 441 AspAsnProArgTyrcLysGlyGlyAspLeuThrLeuTyrcLysIleAsnLeuHleuVal 460
QY 1443 ACCAATGACTGGGCTTACCTTATCTTTTCTAACAAGTGAATAAATCTTCTTA 1502
Db 461 ThrLysIleLeuArgLeuProTyrcLysProPheSerAsnLysGlnAlaAspTyrcLysLeu 480
QY 1503 AGACCTTGGGACCTCATGATGATTAATCTTCCAAATGTCCTCAACTCAATGCTTAATCTTA 1562
Db 481 ArgProLeuGlyProHleuGlyLeuLeuSerLysSerValGlnLeuAsnIleLeuThrLeu 500
QY 1563 AAGATGATGATGATCAAAACCTTGGCCACTTTATGAAAGAAACCTTCCGGCCAGAGAT 1622
Db 501 LysMetValAspAspGlnThrLeuProProLeuMetGlnLysBProLeuArgProGlySer 520
QY 1623 TCATGCGGCTGCGCAGCTTCTCATATAGTTTCTTGTGATAGAAATGCCAAAGTGTGCT 1682
Db 521 SerLeuGlyLeuProAlaPheSerTyrcSerPheValIleArgAsnAlaLysValAla 540
QY 1683 GCTTGCATC 1691
Db 541 AlaCysIle 543

```

RESULT 10
AAV17082
ID AAV17082 standard; protein; 543 AA.
XX
XX AAV17082;

XX 21-JUL-1999 (first entry)
XX
XX
XX

DE Human heparanase enzyme.

KW Heparanase; endoglucuronidase; heparan sulfate proteoglycan; enzyme;
KW metacastis; angiogenesis; wound healing; angiogenesis-induced revascularization;
KW arteriosclerosis; atherosclerosis; inflammation; tissue development;
KW human; HSPG.

OS Homo sapiens.

XX
XX
PN W09921975-A1.

XX 06-MAY-1999.
 PD
 XX 28-OCT-1998; 98MO-AU000898.
 PF
 XX 28-OCT-1997; 97AU-0000062.
 PR 09-DEC-1997; 97AU-00000812.
 XX
 XX (AUSU) UNIV AUSTRALIAN NAT.
 PA
 PI Freeman CG, Hulett MD, Parish CR, Hamdorf BJ;
 XX WPI, 1999-312956/26.
 DR N-PSDB; AAX37259.
 XX
 XX Polynucleotides encoding mammalian endoglucuronidases, especially
 PT heparanases, useful to promote wound healing.
 CC
 PS Claim 6; Page 69-73; 112pp; English.

The invention relates to nucleic acid sequences that encode heparanase enzymes having endoglucuronidase activity. Recombinant heparanases are capable of removing the HS side chain from heparan sulfate proteoglycan (HSPG). Sulfated oligosaccharides, sulphates or HSPG can be used to inhibit heparanase, this is useful for treatment of a physiological or medical condition associated with elevated heparanase activity, such as metastasis, angiogenesis, wound healing, angioplasty-induced restenosis, arteriosclerosis, atherosclerosis and inflammation. The human, murine and rat heparanases can be used to enhance wound healing, especially associated with tissue development and repair. The conditions mentioned above can be diagnosed using specific antibodies, and also using primers and probes specific for the heparanase polynucleotides. Other uses of the heparanases include sequencing sulfated molecules such as HSPG. The present sequence represents a human heparanase

XX Sequence 543 AA;

XX Alignment Scores:

Pred. No.: 3 01e-272 Length: 543
 Score: 2838.00 Matches: 542
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.82% Mismatches: 0
 Query Match: 90.99% Indels: 0
 DB: 2 Gaps: 0

US-10-676-079-3 (1-1721) x AAY17082 (1-543)

QY 63 ATGCTGCTGCGCTCGAAGCTGCGTGCCTGCGCGCTGATGCTGCTGCGGCGG 122
 DB 1 MetLeuLeuArgSerIleProAlaLeuProProLeuMetLeuLeuLeuGlyPro 20
 QY 123 CTGGGCTCCCTCTCCCTGGCGCGCTGCCCACTGCGCAAGCAGAGACTGCGAC 182
 DB 21 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaGlnPValValaIa 40
 QY 183 CTGGAAGCTTTCACCGAGGCGCGTGCACCTGAGGCGCTGCTGCTGCTGCGGAC 242
 DB 41 LeuAspPhePheThrGlnProLeuHisLeuValSerProSerPheLeuSerValThr 60
 QY 243 ATGAGCGCAACCTGCGCAGGACCGCGGTTCTCTATCTCTGGGTTCTCCAAAGTT 302
 DB 61 IleAspAlaAsnLeuAlaThrAspProAlaGlyPheLeuGlySerProGlyLeu 80
 QY 303 CGTACCTTGGCCAGAGGCTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 362
 DB 81 ArgThrLeuAlaArgIleLeuSerProAlaIleLeuArgPheGlyGlyThrIleSerThr 100
 QY 363 TTCCTAATTTTGGATCCCAAGAGAAATCACTTGAAGAGAGAGAACTAGCGCAAT 422
 DB 101 PheLeuIlePheAspProGlySerGlySerThrPheGlnGlnArgSerIleTrpGlnSer 120
 QY 423 CAGTCACACGAGATATTTGCAATATGATCATCTCTGATGTGAGAGAGAACTTA 482

DB 121 GlnValAsnGlnAspIleCysIleYrGlySerIleProProAspValGlnGlnLeu 140
 QY CGGTGGAATGGCCCTACAGAGCAATGCTACTCCGAGAACTACAGAAAAAGTTC 542
 DB 141 ArgLeuGlnTrpProIleYrGlnGlnLeuLeuLeuArgGlnHisIleYrGlnIleAspPhe 160
 QY 543 AAGAACAGACCTACTCAAGAGCTGTAGATGTGCTATACACTTTTGCAAACTGCTCA 602
 DB 161 LysAsnSerThrIleYrSerArgSerSerValAspValLeuIleYrThrPheAlaAsnCysSer 180
 QY 603 GCACTGCACTTGATCTTTGGCTTAAATGCGTTATTAAGACAGCAAGATTGCAATGCGAC 662
 DB 181 GlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGlnTrpAsn 200
 QY 663 AGTCTAATGCTAGTTCCTGCTGACTACCTCTTCAAGGGGATTAACATTTCTGG 722
 DB 201 SerSerAsnAlaGlnLeuLeuLeuAspIleYrCysSerSerIleGlyTrpAsnIleSerTrp 220
 QY 723 GAAGTACGCAATGAACTTACAGATTTCTTAAAGAGCTGATATTTTCATCATGCGTGC 782
 DB 221 GlnLeuGlyAsnGlnProAsnSerPheLeuIleYrValAspIlePheIleAsnGlySer 240
 QY 783 CAGTTAGAGAAATTTATTTCAATTGCAATTAATCTTCAAGAAAGTCCACTTCAAAAT 842
 DB 241 GlnLeuGlyGlnAspPheIleGlnLeuHisIleYrSerLeuAspGlySerThrPheIleAsn 260
 QY 843 GCAAACTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 902
 DB 261 AlaIleLeuIleYrGlyProAspValGlyGlnProArgGlySerThrAlaIleMetLeuIle 280
 QY 903 AGCTTCTGAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 962
 DB 281 SerPheLeuIleValIleGlyGlyValIleAspSerValThrPheIleHisIleYrIleLeu 300
 QY 963 AATGAGCACTGCTACTACAGAGAAATTTCTAAACCTGATGATTTGAATTTTATT 1022
 DB 301 AsnGlyArgThrAlaThrArgIleAspPheLeuAsnProAspValLeuAspIlePheIle 320
 QY 1023 TCATCTGCAAAAAAGTTTTCAGGTGCTGAGAGACAGGCGCTGCAAGAGTCTGG 1082
 DB 321 SerSerValGlnIleValPheGlnAlaValGlnSerThrArgProGlyIleValValTrp 340
 QY 1083 TTAGAGAAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1142
 DB 341 LeuGlyGlnThrSerSerIleIleYrGlyGlyAlaPheLeuLeuSerAspThrPheIle 360
 QY 1143 GCTGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1202
 DB 361 AlaGlyPheMetTrpLeuAspIleLeuGlyLeuSerIleAlaArgMetGlyIleGlnValVal 380
 QY 1203 ATGAGGCAAGTATCTTTGAGAGAGAGAACTACATTTAGTGAGTGAATGAACTTGA 1262
 DB 381 MetArgGlnValPhePheGlyAlaGlyAsnIleValIleValIleAspGlnAspPheAsp 400
 QY 1263 TTACCTGATTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1322
 DB 401 LeuProAspIleYrTrpLeuSerIleLeuPheIleValIleGlyThrIleValLeuMet 420
 QY 1323 GCAAGCGTCAAGGTTTCAAGAGAAAGAACTTGCAGTATGCTTCAATGCAAAACAT 1382
 DB 421 AlaSerValGlnGlySerIleValArgIleGlyLeuValIleGlyThrIleAsnThr 440
 QY 1383 GACATTCGAAGTATTAAGAGAGATTTAACTGTATGCTATTAACCTCAATACCTC 1442
 DB 441 AspAsnProArgIleYrGlyGlnGlyAspLeuThrLeuIleValIleAsnIleHisVal 460
 QY 1443 ACCAAGTCTTGGCGTTACCTTCTTCTTCTTCAAGCAAGTGGATTAATACCTTCT 1502
 DB 461 ThrIleYrIleValGlnLeuProIleYrProPheSerAsnIleGlnValAspIleYrIleLeu 480
 QY 1503 AGACCTTGGAGCTCATGATTTACTTTCAAATGCTGCAATGCTCAATGCTTAACTTA 1562
 DB 481 ArgProLeuGlyProIleGlyLeuLeuSerIleValGlnLeuAsnGlyLeuThrIle 500

QY 1563 AAGATGGTGAATGATCAACCTTGGCCACTTTTAATGAAAAAAGCTTCGCCGCAAGAACT 16722
DB 501 LysMetValaAspArgInThrLeuProLeuMetGluYsProLeuArgProGlySer 520
QY 123 TCACCTGGCTGGCCAGCTTTTCATATGTTTTTTTGATAGAAAAATGCCAAAGTTGCT 1682
DB 521 SerLeuGlyLeuProAlaPheSerTyrSerPheValIleArgAsnAlaValAla 540
QY 1683 GCTTGATC 1691
DB 541 AlaCysIle 543
RESULT 11
AAB86206
ID AAB86206 standard; protein; 543 AA.
AC AAB86206;
XX 24-AUG-2001 (first entry)
DT Human heparanase inhibitor protein.
DE Heparanase; inhibitor; cardiac insufficiency; cardiatic; nephrotropic;
KM hepatotropic; veterinary medicine; congestive heart failure; dyspnoea;
KW primary cardiomyopathy; peripheral edema; pulmonary congestion;
KM hepatic congestion; hydrothorax; ascites; nocturia; human.
XX Homo sapiens.
OS
XX DE19955803-A1.
PN 23-MAY-2001.
XX 19-NOV-1999; 99DE-01055803.
PF 19-NOV-1999; 99DE-01055803.
XX 19-NOV-1999; 99DE-01055803.
PR (KNOL) KNOLL AG.
XX Heir D, Hahn A, Laux V;
XX WPI: 2001-368371/39.
DR N-PSDB; AAH20940.
XX Treatment or prevention of cardiac insufficiency and related conditions,
PT e.g. pulmonary congestion and dyspnoea, comprises administration of
PT heparanase inhibitor.
XX
PS Disclosure; Page 11-13; 16pp; German.
XX This invention describes a novel heparanase inhibitor which can be used
CC for the treatment or prevention of cardiac insufficiency and associated
CC indications, symptoms and/or malfunctions. The heparanase inhibitor of
CC the invention has cardiant, nephrotropic and hepatotropic activity. The
CC products of the invention can be used in human and veterinary medicine,
CC for the treatment or prevention of congestive heart failure e.g. primary
CC cardiomyopathy. Associated conditions treated or prevented with the
CC inhibitor are especially peripheral edemas, pulmonary and hepatic
CC congestion, dyspnoea, hydrothorax and ascites. Renal problems, e.g.
CC nocturia can also be treated. This sequence represents the human
CC heparanase protein described in the method of the invention
XX
SQ Sequence 543 AA;
Alignment Scores:
Pred. No.: 3.01e-272 Length: 543
Score: 2838.00 Matches: 542
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.82% Mismatches: 0
Query Match: 90.99% Indels: 0
DB: 4 Gaps: 0

US-10-676-079-3 (1-1721) x AAB86206 (1-543)
QY 63 ATCTGTGGCTGCGAAGCGCTGGCGCGCGCTGATGCTGCTGCTGGGCGG 122
DB 1 MetLeuLeuArgSerIysProAlaLeuProProProLeuMetLeuLeuLeuGlyPro 20
QY 123 CTGGGTCCTCTCTCCCTGGCGCGCTGGCGCGCAAGCAAGCAAGCTGCGAC 182
DB 21 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaGlnAlaValaAsp 40
QY 183 CTGAACTTTTTCACCCAGAGCGCTGCACTGTGAGCCCTGCTGCTGCTGCTGAC 242
DB 41 LeuAspPhePheThrGlnGlnProLeuHISLeuValSerProSerPheLeuSerValThr 60
QY 243 ATTAGCGCAAGCTGGCAAGCAAGCGCGGTTCTCATCTCTGCGTTCTCCAAAGCTT 302
DB 61 IleAspAlaAsnLeuAlaThrAspProArgPheLeuIleLeuLeuGlySerProIysLeu 80
QY 303 CGTAACTTGGCCAGAGCTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 362
DB 81 ArgThrLeuAlaArgGlyLeuSerProAlaTyrLeuArgPheGlyGlyThrIysThrAsp 100
QY 363 TTCCTAATTTTTCATCCCAAGAAAGATCAACCTTTGAAGAGAGATTAAGGCAATCT 422
DB 101 PheLeuIlePheAspProIysGlySerThrPheGlnGlnArgSerTyrTrpGlnSer 120
QY 423 CAAGTCAACCGAGATATTTGCAATATGATATCCATCCCTGATGATGAGAGAAAGTTA 482
DB 121 GlnValAsnGlnAspIleCysIysTyrGlySerIleProProAspValGlnGlnIysLeu 140
QY 483 CGGTGGAAATGGCCCTTACCAAGAGCAATGCTATCTCCGAAACATCAAGAAAGTTG 542
DB 141 ArgLeuGlnItrProIyrgInGlnGlnLeuLeuArgGlnIleTyrGlnIysIysPhe 160
QY 543 AAGAACAGCACTACTCAAGAAAGCTGTAGATGCTATACACTTTGCAACAGCTCA 602
DB 161 LysAsnSerThrTyrSerIysSerValaAspValLeuTyrThrPheAlaAsnCySer 180
QY 603 GGACTGACTGATCTTTGGCCCTAAATGCGTTATTAAGAACAGCAAGATTGCAAGTGAAC 662
DB 181 GlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGlnTrpAsn 200
QY 663 AGTTCTAATGCTCAAGTTGCTCTGCACTACTGCTCTTCCAAAGGGGATTAACATTCTGG 722
DB 201 SerSerAsnAlaGlnLeuLeuLeuAspTyrCysSerSerIysGlyTyrAsnIleSerTrp 220
QY 723 GAACCTGAGCAAGCACTAAGCTTCTCTTAAGAGGCTGATTTTTCATCAAGGGGTG 782
DB 221 GlnLeuGlyAsnGlnItrProAsnSerPheLeuIysValaAspIlePheIleAsnGlySer 240
QY 783 CAGTTAGAGAAAGATTATATTCATATTCATTAACCTTCAAGAAAGTCCACTTCAAAAT 842
DB 241 GlnLeuGlyGlnAspPheIleGlnLeuHISLysLeuLeuArgLysSerThrPheIysAsn 260
QY 843 GCAAAACCTTAAGTGTCTGATGTTGTGCTCAGCTCCGAAAGAAAGCGCTTAAGTCTGAAG 902
DB 261 AlaLysLeuTyrGlyProAspValGlyGlnProArgArgIysThrAlaLysMetLeuIys 280
QY 903 AGCTTCTGAAGCGTGTGAGAGAGATTCATGATTCATGAGCAATGCAATCAATTG 962
DB 281 SerPheLeuIysAlaGlyIysGlnValIleAspSerValThrTrpHisTyrIysLeu 300
QY 963 AATGACGAGCTGCTACAGGAGAAAGATTCTTAACCTTGATGTATTTGACATTTTAT 1022
DB 301 AsnGlyArgThrAlaThrArgGlnAspPheLeuAsnProAspValLeuAspIlePheIle 320
QY 1023 TCATCTGTGCAAAAAGTTTTCAGGTGTTGAGAGACCAAGCTGGCAAGAGCTCTGG 1082
DB 321 SerSerValGlnIysValPheGlnValValGlnSerThrArgProGlyIysIysValTrp 340
QY 1083 TTAGAGAAACAAGCTGCTCAATGAGAGCGGAGGCGCTTCTCTTCCGACACTTTGGA 1142
DB 341 LeuGlyGlnThrSerSerAlaTyrGlyIysGlyAlaProLeuLeuSerAspThrPheAla 360

QY 1143 GCTGGCTTATATGCTGGATAAATTGGCCCTGTCAAGCCGATGGGAATAGAGTGCG 1202
 DB 361 AAGAGPhenMetTTPLeuaspLysLeuGlyLeuSerAlaArgMetClyLleGluValVal 380
 QY 1203 ATGAGGCAAGTATTTCTTGGAGCAGAAATACCATTTAGTGAGTAAACTTCATCCT 1262
 DB 381 MetArgGlnValPhePheGlyAlaGlyAenTYHLeuValAspGluAsnAspMetPro 400
 QY 1263 TTACCGATTATTTGGTATCTCTTCTGTTCAAGAAATGGTGGCCCAAGGTGTTAAG 1322
 DB 401 LeuProAspTYrTTPLeuSerLeuPheLysLeuValGlyThyrsValLeuMet 420
 QY 1323 GCAAGGTGCAAGGTTTCAAGAGAGAAAGACTTCAGTATACCTTCATTCACAAACT 1382
 DB 421 AlaSerValGlnGlySerIleArgValGlyLeuHisCysThrAsnThr 440
 QY 1383 GCAATTCAGAGTATTAAGAGAGATTTAACTCTGTATGCAATAAACCCTCATACGTC 1442
 DB 441 AspaenProArgTYrIleGlyGlnGlyAspLeuThrLeuTYrAlaIleAsnLeuHisAsnVal 460
 QY 1443 ACCAAGTACTTGGGTTACCTTATCCTTTTCTTAACAGCAAGTGAATAATACCTTCTA 1502
 DB 461 ThrLysTYrLeuArgLeuProTYrProPheSerAsnLysGlnValAspLysTYrLeuLeu 480
 QY 1503 AGACCTTTGGGACCTCATGATTAATTCTTCAAAATGTGCAACTCAATGCTTAACCTTA 1562
 DB 481 ArgProLeuGlyProHisGlyLeuLeuSerLysSerValGlnLeuAsnLysLeuThrLeu 500
 QY 1563 AAGATGTGATGATGATCAAACTTGGCCACTTTAATGAAAACTTCCGGCCAGAGACT 1622
 DB 501 LysMetValAspAspGlnThrLeuProPheLeuMetGlnLysProLeuArgProGlySer 520
 QY 1623 TCACCTGGCTTGGCCGCTTTCATTAATTTTGTGATTAAGAAATGCCAAAGTTGCT 1682
 DB 521 SerLeuGlyLeuProAlaPheSerTYrSerPhePheValIleArgAsnAlaLysValAla 540
 QY 1683 GCTTGATC 1691
 DB 541 AlaCysIle 543
 RESULT 12
 ADD18950
 ID ADD18950 standard; protein, 543 AA.
 AC ADD18950;
 AD 15-JAN-2004 (first entry)
 DE Human disease related protein Segid439.
 DT human, disease state; cytostatic; antiinflammatory; ophthalmological;
 DE antiarteriosclerotic; vlnnerary; gene therapy;
 KW hypoxia-regulated condition; tumorigenesis; angiogenesis; apoptosis;
 KW inflammation; erythropoiesis; glycolysis; gluconeogenesis;
 KW glucose transport; catecholamine synthesis; iron transport;
 KW nitric oxide synthesis; cancer; ischaemic condition; reperfusion injury;
 KW retinopathy; neonatal stress; pre-eclampsia; atherosclerosis;
 KW inflammatory condition; wound healing.
 KM Homo sapiens.
 OS WO2003018621-A2.
 PN 06-MAR-2003.
 PD 23-AUG-2002; 2002WO-GB003892.
 PF 23-AUG-2001; 2001GB-00020558.
 PR 05-OCT-2001; 2001GB-00024037.
 PA (OXFO-) OXFORD BIOMEDICA UK LTD.

PI Kingman SW, White J, Ward NR, Harris RA, Naylor S, Mundy CR;
 XX MPI: 2003-290046/28.
 DR N-PSDB; ADD18951.
 XX New substantially purified polypeptide, useful for diagnosing or treating
 PT a hypoxia-regulated condition, such as cancer, ischemia, reperfusion
 PT injury, retinopathy, pre-eclampsia, atherosclerosis, inflammation, or
 PT wound healing.
 XX Claim 25; SEQ ID NO 439; 424pp; English.
 PS
 XX This invention relates to novel human genes and gene product which are
 CC implicated in certain disease states. Compounds which modulate the
 CC proteins of the invention may have cytostatic, antiinflammatory, the
 CC ophthalmological, antiarteriosclerotic or vlnnerary activities. The
 CC sequences of the invention may be useful for gene therapy. The invention
 CC may be useful for diagnosing or treating a hypoxia-regulated condition,
 CC such as tumorigenesis, angiogenesis, apoptosis, inflammation,
 CC erythropoiesis, or the biological response to hypoxia conditions
 CC including processes such as glycolysis, gluconeogenesis, glucose
 CC transport, catecholamine synthesis, iron transport or nitric oxide
 CC synthesis. The disease includes cancer, ischaemic conditions, reperfusion
 CC injury, retinopathy, neonatal stress, pre-eclampsia, atherosclerosis,
 CC inflammatory conditions or wound healing. The present sequence is that of
 CC a disease related protein of the invention.
 SQ Sequence 543 AA;
 Alignment Scores:
 Pred. No.: 3,01e-272 Length: 543
 Score: 2838.00 Matches: 542
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.82% Mismatches: 0
 Query Match: 90.99% Indels: 0
 DB: Gaps: 0
 US-10-676-079-3 (1-1721) x ADD18950 (1-543)
 QY 63 ATGCTGCTGCGCTGCAAGCTGCGCTGCCGCCGCTGATGCTGCTCTGGGCGCG 122
 DB 1 MetLeuLeuArgSerLysProAlaLeuProProPheLeuMetLeuLeuLeuLysPro 20
 QY 123 CTGGGTCCCTCTCCCTCGGCGCCCTGCGCACTGCGCAAGCAGAGAGTGCTGAC 182
 DB 21 LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaAspValAlaAsp 40
 QY 183 CTGACCTTCTTCAACCAGAGCGGCTGCACTGTGAGCCCTCGTCTCTGCTCACC 242
 DB 41 LeuAspPhePheThrGlnGluProLeuHisLeuValSerProSerPheLeuSerValThr 60
 QY 243 ATGAGCGCAACCTGCGCAAGCAGACCGCGGTTCTCTGATCTCTGCTTCTCAAGCTT 302
 DB 61 IleAspAlaAsnLeuAlaThrAspProArgPheLeuIleLeuGlySerProLysLeu 80
 QY 303 CGTACTTGGCCAGAGCTTGTCTCTCGCTACCTGAGTTTGTGGGACCAAGCAAGC 362
 DB 81 ArgThrLeuAlaArgGlyLeuSerProAlaTYrLeuAspPheGlyGlyThyrsThrAsp 100
 QY 363 TTCTTAATTTTGAATCCCAAGAGAGAAATCAACTTTGAAGAGAGAGTAACTGCAATCT 422
 DB 101 PheLeuIlePheAspProLysLysGluSerThrPheGlnGlnLysSerTYrTTPGlnSer 120
 QY 423 CAAGTCAACAGATATTTGCAATAATATGATTCATCCCTCTGATGTGAGAGAGATT 482
 DB 121 GlnValAlaGlnAspLysCysLysTYrGlySerIleProProAspValGlnGlnLysLeu 140
 QY 483 CGGTGGAATGGCCCTTACAGAGCAATTTGCTACTCGAGAACTCAACAAAAGTTG 542
 DB 141 ArgLeuGlnTTPProTYrGlnGlnGlnLeuLeuLeuAspGlnHisTYrGlnLysLysPhe 160
 QY 543 AAGAACAGACCTACTCAAGAGAGCTGTGATGTGCTATACATTTTGAACAGTCTCA 602

```

Db      161  LysAsnSerThrTyrSerArgSerSerValAspValLeuTyrThrPheAlaAsnCysSer 180
Qy      603  GGAAGTGAATGATCTTGGCTTAATGCGTTATTAAAGAACAGAGATTTGGAGTGGAGAC 662
Db      181  G1LeuAspLeuLeuPhePheG1LeuAsnAlaLeuLeuArgThrAlaAspLeuGlnTyrAsn 200
Qy      663  AGTTCTAATGCTCAGTTGCTCTGAGTCTGCTTCCAAAGGGGATTAACATTTCTTGG 722
Db      201  SerSerAsnAlaGlnLeuLeuAspTyrCysSerSerTyrGlyTyrAsnIleSerTyr 220
Qy      723  GAATAGAGCAATGAACCTAACAGTTTCCTTAAGAAAGGCTGATATTTCATCAATAGGGTGG 782
Db      221  GlnLeuG1LysGlnTyrProAsnSerPheLeuLysValAspIlePheIleAsnGlySer 240
Qy      783  CAGTTAGAGAAAGATTTATTTCAATTGTCATTAACCTTCAAGAAAGTCCACTTCAAAAAT 842
Db      241  GlnLeuG1LysGlnTyrProAsnSerPheLeuLysValAspIlePheIleAsnGlySer 260
Qy      843  GCAAAATCTAATGCTCAGTTGCTCTGAGTCTGCTTCCAAAGGGGATTAACATTTCTTGG 902
Db      261  AlAluLeuTyrG1LysProAspValG1GlnProArgTyrG1ThrAlaLysMetLeuLys 280
Qy      903  AGGTCCTGAGAGGCTGGAGAGAGATTTGATTCAGTTACATGAGCTCACTACTATTGG 962
Db      281  SerPheLeuLysAlaG1LysG1LysValIleAspSerValThrPheIleAsnGlySer 300
Qy      963  AATGAGCGAGCTGCTACAGGAGAGATTTTCTAAACCTGATGATTTGACATTTTATTT 1022
Db      301  AsnG1LysGlnTyrAlaThrArgG1LysAspPheLeuAsnProAspValLeuAspIlePheIle 320
Qy      1023  TCATCTGTGCAAAAAGTTTTCAGAGTGTGAGAGCAAGGCTGGCAAGAGTGTGG 1082
Db      321  SerSerValGlnLysValPheGlnValAlG1LysSerThrArgProG1LysValLysValTyr 340
Qy      1083  TTAGAGAGAAACAAGCTCTGCAATAGAGAGAGAGCGGCTTGCATCCGACACCTTTGCA 1142
Db      341  LeuG1LysGlnThrSerSerAlaTyrG1LysValAlaProLeuLeuSerAspThrPheAla 360
Qy      1143  GCTGGCTTTATGTGGTGAATTAATTGGCCCTGTCAGCCCGAATGGAGATTAAGAGTGTG 1202
Db      361  AlAlG1LysMetCysPheLeuAspLysLeuG1LysSerAlaArgMetG1LysGlnValVal 380
Qy      1203  ATGAGGCAATGATTTCTTGGAGAGAGAGAACTACATTTAGTGGATGAAAACCTTGATCCT 1262
Db      381  MetArgGlnValPhePheG1LysAlaG1LysAsnTyrIleLeuValAspG1LysAsnPro 400
Qy      1263  TTACCTGATTTATGTGGTGAATTTGAGAGAAATTTGGGGGCAACAAGTGTATATG 1322
Db      401  LeuProAspTyrTyrPheLeuSerLeuPheLysLeuValG1LysThrValLeuMet 420
Qy      1323  GCAAGCGTGAAGGTTCAAGAGAGAGAGAGCTTGAATATACCTTCAATGCAACAACCT 1382
Db      421  AlAserValGlnG1LysSerLysArgArgLysLeuArgValTyrLeuHisCysThrAsnThr 440
Qy      1383  GACAAATCCAAAGGTTTAAAGAGAGATTTAACTGTATGTCATTAACCTTCATTAAGTTC 1442
Db      441  AspAsnProArgTyrLysGlnG1LysAspLeuThrLeuTyrAlaIleAsnLeuHisAsnVal 460
Qy      1443  ACCAAGACTTGGCGGTTACCTATCCTTTTCTAAAGAGAGAGTGAATTAATACCTTCTA 1502
Db      461  ThrLysTyrLeuAspGlnLeuProTyrPhePheSerAsnLysGlnValAspLysTyrLysLeu 480
Qy      1503  AGACCTTTGGGAGCTCATGATTAATCTTCCAAATCTGTCCAACTCAATGGTCTTAACCTTA 1562
Db      481  ArgProLeuG1LysProHisG1LysLeuLeuSerLysSerValGlnLeuAsnG1LysLeuThrLeu 500
Qy      1563  AAGATGTTGATGATTAACCTTTCACCTTTTAAATGAAAAAAGCTCTCCGGCCAGAGAGT 1622
Db      501  LysMetValAspAspGlnThrProLeuMetG1LysProLeuAspProG1LysSer 520
Qy      1623  TCACCTGGGCTGGCAGGCTTCTCATATAGTTTCTTGTGATTAAGAAATGGCAAGTGTCT 1682
Db      521  SerLeuG1LysLeuProAlaPheSerTyrSerPhePheValIleArgAsnAlaLysValAla 540

```

```

Qy      1683  GCTTGCATC 1691
Db      541  AlaCysIle 543
RESULT 13
ID AAB88361 standard; protein, 543 AA.
XX
AC AAB88361;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human membrane or secretory protein clone PSEC0090.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;
KM rheumatoid arthritis; diabetes.
XX
OS Homo sapiens.
XX
PN EP1067182-A2.
XX
PD 10-JAN-2001.
XX
PF 07-JUL-2000; 2000EP-00114090.
XX
PR 08-JUL-1999; 99JP-00194179.
PR 11-JAN-2000; 2000JP-00118775.
PR 02-MAY-2000; 2000JP-00183766.
XX
PA (HELI-) HELIX RES INSTR.
XX
PI Oka T, Iwogai T, Mishioka T, Kawai Y, Sugiyama T, Hayashi K;
DR WPI; 2001-093989/11.
DR N-PSDB; AAF93788.
XX
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
PS gene therapy or as candidate target molecules in drug development.
XX
PS Claim 1; SEQ ID NO 90; 609pp + Sequence listing; English.
XX
CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by AAB88317
CC - AAB88412. Included in the invention are primers AAF93917 - AAF94235 and
CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the
CC invention. The invention also includes methods for the production of
CC antibodies directed against the proteins, and cDNA sequences, which can
CC be used in vaccines. The polynucleotide sequences can be used in gene
CC therapy. The polynucleotide sequences and the proteins they encode may be
CC used in the prevention, treatment and diagnosis of diseases associated
CC with inappropriate secretory protein/membrane protein expression. The
CC nucleic acids and complementary sequences may also be used as DNA probes
CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
CC and quantitate the presence of similar nucleic acid sequences in samples.
CC They may also be used to study the expression and function of secretory
CC proteins/membrane polypeptides and their role in metabolism. The
CC polypeptides may be used as antigens in the production of antibodies
CC against them and in assays to identify modulators (agonists and
CC antagonists) of expression and activity. The antibodies and antagonists
CC may also be used as therapeutic agents to down regulate expression and
CC activity. The antibodies may also be used as diagnostic agents for
CC detecting the presence of the polypeptides in samples (e.g. by enzyme
CC linked immunosorbent assay (ELISA)). Examples of diseases which may be
CC treated include rheumatoid arthritis and diabetes
XX
SQ Sequence 543 AA;

```

Alignment Scores:

Pred. No.:	4,68e-271	Length:	543
Score:	2826.00	Matches:	540
Percent Similarity:	99.82%	Conservative:	2
Best Local Similarity:	99.45%	Mismatches:	1

Query Match:	90.61%	Indels:	0
DB:	4	Gaps:	0

US-10-676-079-3 (1-1721) × AAB88361 (1-53)

QY	63	ATGCTGCTGGCCTGGAAAGCTTGGCCGCTGGCCGCGCGCTGAATGCTGCTCTCTGGGGCCG	122
Db	1	MetLeuLeuAspSerLysProAlaLeuProProProLeuMetLeuLeuLeuGlyPro	20
QY	123	CTGGATCCCTCTCCCTCGGCGCCCTGGCCGCACTGGCGCAAGACAGAGCGTGGAAC	182
Db	21	LeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaGlnAspValAsp	40
QY	183	CTGGACTTCTTCAACCAGAGACCCTGCACTGGTAGCCCTCTGCTCTGTCGCTCAC	242
Db	41	LeuAspPhePheThrGlnGluProLeuHisLeuValSerProSerPheLeuSerValThr	60
QY	243	ATTAAACGCAACTGGGCGACGGACCCGGGGTTCCTCATCTCCCTGGGTTCCCAAGCTT	302
Db	61	IleAspAlaAsnLeuAlaThrAspProArgPheLeuIleLeuLeuGlySerProLysLeu	80
QY	303	CGTACTTGGCCAGAGCGCTTGCTCTCTCGCTACCTGAGTTTGGTGCAACCAAGAC	362
Db	81	ArgThrIleuAlaArgGlyLeuSerProAlaTyrIleuArgPheGlyGlyThrLysThrAsp	100
QY	363	TTCTTAATTTTCGATCCCAAGAAGAAATCAACTTTGAAGAGAGAATTACGGCAATCT	422
Db	101	PheLeuIlePheAspProLysLeuGlySerThrPheGluGluArgSerTyrTrpGlnSer	120
QY	423	CAAGTCACACAGAGATATTGGCAAAATATGATGCATCCCTCGTGAATGTGAGAGAAAGTTA	482
Db	121	GlnValaAsnGlnAspIleCysLysTyrIolysertIleProPheAspValGluGluLysLeu	140
QY	483	CGGTTGGAATGACCCCTTACCAAGAGCAATGTGCTACTCCGAGAACATCACAGAAAAAGTTC	542
Db	141	ArgLeuGluIuTrpProTyrGlnGluGlnLeuLeuAspGlyIuHisTyrGlnLysPhe	160
QY	543	AAGAACAGCACTCTCTCAAGAGCGTCTGTATGCTATACCTTTTGGCAACATGCTCA	602
Db	161	LysAsnSerThrTyrSerArgSerSerValAspValLeuTyrThrPheAlaAsnCysSer	180
QY	603	GGACTGCACTTGATCTTTGGCCTTAATCGTTATTAAACAAGCAGATTTCAGTGAAC	662
Db	181	GlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuGlyThrAlaAspLeuGlnTrpAsn	200
QY	663	AGTTCTAATGCTCACTGCTCTCTGGACTACTGCTTCTCAAGGGGTATAACATTTCTTGG	722
Db	201	SerSerAsnAlaGlnLeuLeuLeuAspTyrCysSerSerLysGlyTyrAsnIleSerTrp	220
QY	723	GAACTAAGCAATGAACTCAACAGTTTCTTAAAGAGGTGATATTTTACATAAAGGGTCG	782
Db	221	GluLeuGlyAsnGluProAsnSerPheLeuLysLysAlaAspIlePheIleAsnGlySer	240
QY	783	CAGTTAGAGAGAATTATTAATTCAATTGATAAACTTTAAGAAAGTCCACCTTCAAAAT	842
Db	241	GlnLeuGlyGluAspPheIleGlnLeuHisLysLeuLeuArgLysSerThrPheLysAsn	260
QY	843	GCAAAACCTCTATGCTCTCGATGTTGGTCAGCGCTCGAAGAAAGCGGCTTAAGTGTGAAG	902
Db	261	AlaLysLeuTyrGlyProAspValGlyGlnProArgLysThrAlaLysMetLeuLys	280
QY	903	AGCTTCTCGAAGGCGTGTGAGAGATGATGATTAGTTAATGAGCATCAGTACATATTG	962
Db	281	SerPheLeuLysAlaGlyGlyGluValIleAspSerValThrTrpHisIleTyrTrpLeu	300
QY	963	AATGACGGAATGCTTACCAAGGAGATTTTCTAAACCTGTATGTAATTGACATTTTATTT	1022
Db	301	AsnGlyArgThrAlaThrArgLysAspPheLeuAsnProAspValIleuAspIlePheIle	320
QY	1023	TCACTGGTGCAAAAAGTTTCCAGGTGGTTGAGACACAAGCCCTGGCAAAAGTCTGG	1082
Db	321	SerSerValGlnLysValPheGlnValValGluSerThrArgProGlyLysLysValAlaTrp	340

QY	1083	TTAGGAGAAACAACTCTGCATATAGAGAGCGGACGCCCTTGCTATTCGACACTTTTGCA	1142
Db	341	LeuGIgIuItrISerSerIaInIGIyGIyAlaProIeuIeuSerAspThrPheAla	360
QY	1143	GCTGGCTTATAGTGGCTGGAATTAATTGGGCGCTGTCAGCGCCGGAATGGGAATGAAGTGGT	1202
Db	361	AlaGlyPheMetCttrPLeuAspIyIeuGIyIeuSerIaIatGMeGtIyIleGIuValI	380
QY	1203	ATGAGGCAAGATATTCTTTGGAGCAGAAACTACCATTTAGTGATGAAAATTGGATCCT	1262
Db	381	MetArgInValPhePheGIyAlaGIyAsnTyrIleuValaArgIuIaAsnPheAspPro	400
QY	1263	TTACCTGATTATTGGCTATCTCTCTGTTCACAAATAATGGTGGCCACCAAGGTATTATG	1322
Db	401	LeuProAspTyrItrPLeuSerIeuIeuPheIeuIyIeuValIGtYtrIlyValIeuMet	420
QY	1323	GCAAGCGTCGAAGGTTCAAAAGAAAGGAAGCTTGAGATACCTTCATTGACAAAACACT	1382
Db	421	AlaSerValGIInGIySerIyAsnArgIyIeuArgIyIeuValIyIyIeuHisCysThrAsnThr	440
QY	1383	GACAATCCAAAGTATTAAGAAGAGATTTAACCTGTATGSCATTAACCTTCATTAAGTGC	1442
Db	441	AspAsnProAspGtYrIyIyAsnGIyAlaAspIeuThrIeuTyrIaIaIeAsnIeuHisAsnVal	460
QY	1443	ACCAAGTATCTTGCGGTTACCTCATCTCTTTTCTAACAGCAAGTGATTAATACCTTCTA	1502
Db	461	ThrIySyrIyIeuAlaArgIeuProIyProPheSerAsnIySgnIValaAspIySyrIyIeu	480
QY	1503	AGACCTTTGGGACCTGCATGATTACTTTCCAAATCTGTCCAACTCAATGGCTTAACCTTA	1562
Db	481	ArgProIeuGIyProIHaSgIyIeuIeuSerIySerValGIInIeuAsnGIyIeuThrIeu	500
QY	1563	AAGATGTGTGATGATCAAACTTGCCCACTTTAATGGAAGAACTCTCCGGCCAGAAAGT	1622
Db	501	LysMetValaAspAspArgInThrIeuProIeuMetGIuIySPrIeuAsnArgProGIySer	520
QY	1623	TCACCTGGCTTGCCCACTTCTCATATAGTTTTTTGTGATGAAGAAATGCCAAATGTTGCT	1682
Db	521	SerIeuGIyIeuProIaPheSerIySerPheIeuValIleArgAsnIaIyValaIa	540
QY	1683	GCTTGCAATC 1691	
Db	541	AlaCysIle 543	
RESULT 14			
ABP56822			
ID	ABP56822 standard; protein; 545 AA.		
XX	ABP56822;		
AC			
XX			
DT	02-APR-2003 (first entry)		
XX			
DE	Human heparanase protein SEQ ID NO:18.		
XX			
KW	Human; heparanase; phosphorothioate; antisense oligonucleotide;		
KM	cytostatic; gene therapy; tumour; enzyme.		
OS	Homo sapiens.		
XX			
FN	WO2003004705-A1.		
XX			
PD	16-JAN-2003.		
XX			
PF	01-JUL-2002; 2002WO-US020636.		
XX			
PR	05-JUL-2001; 2001US-00899440.		
XX			
PA	(UNYCO) UNIV COLUMBIA NEW YORK.		
XX			
PI	Stein C;		
XX			
WPI	2003-201558/19.		
DR	N-PSDB; AB222816.		

XX New oligonucleotide having a sequence complementary to a sequence of
PT ribonucleic acid encoding a heparanase, useful for preparing a
PT composition for treating tumor.
XX
PS Disclosure; Page 46-47; 48pp; English.
XX The present invention describes an oligonucleotide having a sequence
CC complementary to a sequence of ribonucleic acid encoding a heparanase.
CC The oligonucleotide hybridizes with the ribonucleic acid under conditions
CC of high stringency and has a sequence comprising 10-40 bp. The
CC intermolecular linkages of the oligonucleotide comprise at least one
CC phosphorothioate linkage. Hybridisation of the oligonucleotide to the
CC ribonucleic acid inhibits expression of the heparanase, where inhibition
CC of heparanase means at least a 50% reduction in the quality of
CC heparanase. Also described: (1) a method of inhibiting expression of a
CC oligonucleotide in an amount effective to inhibit the expression of a
CC heparanase in the cell and a carrier; and (3) a method of treating a
CC tumour in a subject comprising administering to the subject an amount of
CC the above oligonucleotide effective to inhibit expression of a heparanase
CC in the subject. Heparanase antisense oligonucleotides have cytostatic
CC activity, can be used in gene therapy, and can be used for preparing a
CC composition for treating tumours. The present sequence represents human
CC heparanase, which is given in the exemplification of the present
CC invention
XX
XX Sequence 545 AA:
SQ
Alignment Scores:
Pred. No.: 3,676-270 Length: 545
Score: 2817.00 Matches: 542
Percent Similarity: 99.63% Conservative: 1
Best Local Similarity: 99.45% Mismatches: 0
Query Match: 90.32% Indels: 2
DB: 6 Gaps: 2
US-10-676-079-3 (1-1721) x ABP56822 (1-545)
QY 63 ATGCTGCTGCGCTGGAAGCCCTGCGCCGCGC---CTGATGCTGCTGCTGCGG 119
Db 1 MetLeuLeuArgSerLysProAlaLeuProProProLeuLeuLeuLeuGly 20
QY 120 CCGGTGGGTCCCTCTCCCTCGGCGCCCTGCGCCGACCTGGCAGCA---CAGAGCTC 176
Db 21 ProLeuGlyProLeuSerProGlyAlaLeuProArgProAlaGlnAlaGlnAspVal 40
QY 177 GTGACCTGGAAGCTTCTCACCGACGACCGCTGACCTGGAGCCCTCGTCTCTCC 236
Db 41 ValAspLeuAspPhePheThrGlnGluProLeuHisLeuValSerProSerPheLeuSer 60
QY 237 GTCAACATTGAAGCAAGCTGGCCAGGACCGCGGGTTCCTCATCTCTGGGTTCCA 296
Db 61 ValThrIleAspAlaAsnLeuAlaThrAspProArgPheLeuIleLeuLeuGlySerPro 80
QY 297 AAGCTTGTAAGCTTGGCCAGAGGCTTGTCTCCGCGTAACTGAGAGTTGGGACCAAG 356
Db 81 LysLeuArgThrLeuAlaArgGlyLeuSerProAlaThrLeuAspPheGlyGlyThrLys 100
QY 357 ACAGACTTCTCAATTTTCGATCCCAAGAGGAATCAACTTGAAGAGAGAGTTACTCG 416
Db 101 ThrAspPheLeuIlePheAspProLysGlySerThrPheGlnGlnArgSerLysTrp 120
QY 417 CAATTCGAAGTCAACAGATATTGGAAATATGATCCATCCCTCTGATGTGGAGAG 476
Db 121 GlnSerGlnValAsnGlnAspIleCysLysGlyGlySerIleProProAspValGlnGlu 140
QY 477 AAGTTACGGTTGGAAATGCGCTTACAGAGCAATTTGCTACTCCGAGAACTACAGAA 536
Db 141 LysLeuArgLeuGlnTrpProLysGlnGlnGlnLeuLeuLeuArgGlnHisGlyGlnLys 160
QY 537 AAGTTCAAGACAGACACTACTCAAGAAAGCTGTAGATGTGCATACCTTTTGAAC 596
|||||

Db 161 LysPheLysAsnSerThrTyrSerArgSerSerValAspValLeuTyrThrPheAlaAsn 180
QY 597 TGCTCAGACTGGAATCTTGTGGCTTAATGCGTTATTAAAGACAGAGATTTGCAG 656
Db 181 CysSerGlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGln 200
QY 657 TGGAAAGTTCTTAATGCTCAGTTGCTCTGAGACTACTGCTTTCGAAAGGGATTAATT 716
Db 201 TrpAsnSerSerAsnAlaGlnLeuLeuLeuAspTyrCysSerSerLysGlyTrpAsnIle 220
QY 717 TCTGGGAATPAGCAATGAAGCTTAACAGTTTCTTAAGAGGCTGATATTTTCATCAAT 776
Db 221 SerTrpGlnLeuGlnAsnGlnProAsnSerPheLeuLysValAlaAspIlePheIleAsn 240
QY 777 GGGTCCAGCTTGAAGAGATTAATATTCATATTCATTAACCTTGAAGAAAGTCCACTTC 836
Db 241 GlySerGlnLeuGlyLysAspPheIleGlnLeuHisLysLeuLeuArgLysSerThrPhe 260
QY 837 AAAAAATGCAAAATCTATGCTCTGATGTGTTGGTTCAGCTTGCAGAAAGACGCTTAAGATG 896
Db 261 LysAsnAlaLysLeuTyrGlyProAspValGlyGlnProArgArgLysThrAlaLysMet 280
QY 897 CTGAAGAGCTTCTGAAGGCTGTGGAGAGAGATTTGATTCAGTTACATGACATCACTAC 956
Db 281 LeuLysSerPheLeuLysAlaGlyGlyGlnValIleAspSerValThrTrpHisLysTrp 300
QY 957 TATTGAAATGACAGCATGCTTACAGAGAGATTTTCTAAACCTGATGATTGACATT 1016
Db 301 TyrLeuAsnGlyArgThrAlaThrArgGlnAspPheLeuAsnProAspValIleAspIle 320
QY 1017 TTTATTTTCATCTGTGCAAAAAGTTTTCAGAGTGTGTAAGACACAGGCTTGGCAAGAG 1076
Db 321 PheIleSerSerValGlnLysValPheGlnValValGlnSerThrArgProGlyLysLys 340
QY 1077 GTCTGGTTAGAGAAACAAGCTCTGCATATGAGAGGAGCGGCTTGCATCCGACACC 1136
Db 341 ValTrpLeuGlyLysThrSerSerAlaTyrGlyGlyAlaIleProLeuLeuSerAspThr 360
QY 1137 TTTGACAGCTGCTTATATGTGCTGATTAATTTGGGCTGTACCGCGAAATGGGAATAGA 1196
Db 361 PheAlaAlaGlyPheMetTrpLeuAspLysLeuGlyLysSerAlaArgMetGlyIleGln 380
QY 1197 GTGATATAGAGCAATATTTCTTTGGACAGCAAACTTACATTTATGATGAATGAACCTTC 1256
Db 381 ValValMetCysGlnValPhePheGlyAlaGlyAsnTyrHisLeuValAspGluAsnPhe 400
QY 1257 GATCTTTTACCTGATTTATGCTATCTCTTCTGTTCAAGAAATTTGGTGGGACCAAGTG 1316
Db 401 AspProLeuProAspTyrTrpLeuSerLeuLeuPheLysLysValGlyThrLysVal 420
QY 1317 TTAATGGCAAGCTGCAAGGTTCAAAAGAGAGAGCTTGAGTATACCTTCATTGCACA 1376
Db 421 LeuMetAlaSerValGlnGlySerLysArgAspGlyLeuAspValTyrLeuHisCysThr 440
QY 1377 AACACTGACAATCCAGATTAATAAGAGAGATTTAACTCTGATATGCCATAACCTCCAT 1436
Db 441 AsnThrAspAsnProArgTyrLysGlnGlyAspLeuThrLeuTyrAlaIleAsnLysHis 460
QY 1437 AACGTCAACAAGTACTGGGTTACCTTATCCCTTTTCTTAACAAGCAAGTGTAAATAC 1496
Db 461 AsnValThrLysTyrLeuArgLeuProTyrProPheSerAsnLysGlnValAspLysTrp 480
QY 1497 CTTCTAAGACTTTGGGACCTCATGGATTACTTCCAAATCTGTCCAACTCAATGGTCTTA 1556
Db 481 LeuLeuArgProLeuGlyProHisGlyLeuLeuSerLysSerValGlnLeuAsnGlnLys 500
QY 1557 ACTCTTAAGATGGTGAATGATCAAACTTGGCCACTTTTAATGGAAAAAAGCTCTCCGGCCA 1616
Db 501 ThrLeuLysMetValAspAspGlnThrLeuProProLeuMetLysLysProLeuAspPro 520
QY 1617 GGAAGTTCACTGGGCTTGGCAAGCTTCTCATATAGTTTGTGATTAAGAAATGGCCAAA 1676
Db 521 GlySerSerLeuGlyLeuProAlaPheSerTyrSerPhePheValIleArgAsnAlaLys 540

Db 161 LysPheLeuAsnSerThrTyrSerArgSerSerValAspValLeuTyrThrPheAlaAsn 180
QY 597 TGCTCAGACGTGACCTTGATCTTTGGCCCTAAATCGTTATTAGACAGCAGATTTTCAG 656
Db 181 CysSerGlyLeuAspLeuIlePheGlyLeuAsnAlaLeuLeuArgThrAlaAspLeuGln 200
QY 657 TGGACAGTTCTAAATGCTCAGTTCCTCTGACACTACCTGCTTTCCAAAGGGGTATTAACATT 716
Db 201 TrpAsnSerSerAsnAlaGlnLeuLeuAsnProTyrCysSerSerLeuGlyTyrAsnIle 220
QY 717 TCTTGGGAACCTAGGCAATGAACCTAACAGTTTCCTTAAAGAGGCTATATTTTCATCAAT 776
Db 221 SerTrpIleuLeuGlyAsnGluProAsnSerPheLeuLeuAlaAspIlePheIleAsn 240
QY 777 GGGTCGACGTTAGAGAGATTAATATTCATTTGCATTAACCTTCTAGAAAGTCCACCTTC 836
Db 241 GlySerGlnLeuGlyIleAspPheIleGlnLeuHisGlyLeuLeuArgGlySerThrPhe 260
QY 837 AAAAAATGCAAACTCTATGCTCTGATGTTGGTCAAGCTCGAAGAAAGCGCTTAAGATG 896
Db 261 LysAsnAlaLeuLeuTyrGlyProAspValGlyIleProArgGlyThrAlaLysMet 280
QY 897 CTGAAGAGCTTCTGAAGGCTGTGAGAGATGATGATTCAGTTACAGTTACATGCGCATCACTAC 956
Db 281 LeuLysSerPheLeuLysAlaGlyGlyGluValIleAspSerValThrTrpHisIleTyr 300
QY 957 TAATTGAATGACGCGACTGTACACGAGAAAGATTTCTAAACCTGATGTAATTGACATT 1016
Db 301 TyrLeuAsnGlyArgThrAlaThrArgIleAspPheLeuAsnProAspValIleAspIle 320
QY 1017 TTTATTTTCATCTGTGCAAAAAGTTTTCAGAGTGTGAGAGCACACAGCCCTGCGAAGAG 1076
Db 321 PheIleSerSerValGlnLysValPheGlnValAlaGlnSerThrArgProGlyLys 340
QY 1077 GTCTGATTAGAGAAACAGCTCTGACATATGAGAGCGAGCGCCCTTGCTATCCGACACC 1136
Db 341 ValTrpLeuGlyGlnThrSerSerAlaTyrGlyGlyAlaProLeuLeuSerAspThr 360
QY 1137 TTTCGACCTGGCTTTATGTGGCTGGATTAATTTGGGCTGTGACCCGAAATGGAAATGAA 1196
Db 361 PheAlaIleGlyPheMetTrpLeuAspLeuGlyLeuSerAlaArgMetGlyIleGln 380
QY 1197 GTGGTGAAGAGCAAGTATCTTTGAGCAGAGAACTAACCTTTAGTGATGAAAACCTTC 1256
Db 381 ValValMetArgGlnValPhePheGlyAlaGlyAsnTyrHisLeuValAspGluAsnPhe 400
QY 1257 GATCCTTTACCTGATTAATTTGGCTATCTCTTCTGTTCAAGAAATTGTGGGCAACCAAGTG 1316
Db 401 AspProLeuProAspTyrTrpLeuSerLeuLeuPheLysLeuValGlyThrLysVal 420
QY 1317 TTAATGGCAAGCTGCAAGGTTCAAGAGAAAGAAAGCTTCGAGTAACTTCATTTGCACA 1376
Db 421 LeuMetAlaSerValGlnGlySerLysArgArgLysLeuArgValTyrLeuHisCysThr 440
QY 1377 AACACGTACAACTCCAAAGTATTAAGAAGAGATTAACTGTATGCCATTAACCTCCAT 1436
Db 441 AsnThrAspAsnProArgTyrLysGlnGlyAspLeuThrLeuTyrAlaIleAsnLeuHis 460
QY 1437 AACGTCAACAAGTACTTGGGTTACCTTACCTTATCTTTCTAAACAAGAGTGATTAATAC 1496
Db 461 AsnValThrIleTyrTyrLeuAspGluLeuProTyrProPheSerAsnLysGlnValAspLysTyr 480
QY 1497 CTTCTTAAGACCTTTGGACCTCATGATTAATTTCCAAAATCTGTCCAACCTCAATGCTCA 1556
Db 481 LeuLeuArgProLeuGlyProHisGlyLeuLeuSerLysSerValGlnLeuAsnGlyLeu 500
QY 1557 ACTCTAAGATGGTGAATGATCAAACTTGCCACCTTAATGAAAACCTTCGGGCA 1616
Db 501 ThrLeuLysMetValAspAspGlnThrLeuProLeuMetGlnLysProLeuArgPro 520
QY 1617 GGAAGTTCACCTGGGCTTGCCAGCTTTCTCATATAGTTTTTTTGGATAGAAAATGCCAAA 1676
Db 521 GlySerSerLeuGlyLeuProAlaPheSerTyrSerPhePheValIleArgAsnAlaLys 540

QY 1677 GTTGTGCTTGCATC 1691
Db 541 ValAlaAlaCysIle 545

Search completed: August 28, 2004, 07:58:11
Job time : 233 secs


```

QY 168 GAGACAGCTGCGAAGCTTCTTCAAGCCGAGGCGGCTGACCTGGTGAAGCCCTCG 227
   ::::: ||||| ::::: ||||| ::::: |||||
Db 58 LyrThrLeuLeuLeuAerValSerThrLysAsnProValArgThrValAsnGluAsn 77
QY 228 TTCTGTCGCTCAGCATTTGACGCCGACCTGGACCGGCGGCTTCTCATCTCTG 287
   ::::: ||||| ::::: |||||
Db 78 PheLeuSerLeuGlnLeuAerProSerLeuLeuAsp--GlyTrpLeuAerPheLeu 96
QY 288 GATTCTCAGAGCTTCTGACCTTGGCAGAGGCTTGTCTGGGCTGACCTGAGGTTGAT 347
   ||| ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 97 SerSerLysArgLeuValThrLeuAlaArgGlyLeuSerProAlaPheLeuArgPheGly 116
QY 348 GGACCAACAGACAGCTTCTTAATTTTC-----GATCCCAAGAAAGATCAACC 395
   ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 117 GlyLysArgThrAerPheLeuGlnPheGlnAsnLeuArgAsnProAlaLysSerArg--- 135
QY 396 TTGGAAGAGAAAGTTACTGGCAATCTCAAGTCAACAGATATTTGCAAATATGATCC 455
   ||| ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 136 -----GlyGly 137
QY 456 ATCCCTCCTGATGAGAGAGAAAGTTACGGTTGGAATGCCCTACAGAGCAATTGCTA 515
   |||||
Db 138 ProGlyProAsp----- 141
QY 516 CTCGAGAACACTACAGAAAAAGTTCAAGAACAGACCTACAGAGAGCTGTAGAT 575
   ::::: ||||| ::::: |||||
Db 142 -----LyrTrpLeuLysAsnTrpGlu----- 148
QY 576 GTGCTATACACTTTTGCAAACTGCTCAGAGACTGACCTTGATCTTGGCTTAATGCTTA 635
   |||
Db 148 ----- 148
QY 636 TTAAGAACAGACAGATTTGAGTGAAGACAGTTCTAATGCTCAGTTGCTCTGACTACTGC 695
   |||
Db 148 ----- 148
QY 696 TCTTCCAGGGGATTAACATTTCTTGGAACTAAGCAATGAACCTTAACAGTTCCCTTAAG 755
   ::||| ||||| ::|||
Db 149 -----AspGlnProAsnAsnTrpArgThr 156
QY 756 AAGCTGATATTTTCATCAATGCGTGCAGTTAGAGAAAGATTTATATCAATTCATMAA 815
   ::||| ||||| ::||| |||||
Db 157 MetHisGlyArgAlaValAsnGlySerGlnLeuGlyLysAspTrpTrpGlnLeuLysSer 176
QY 816 CTTCATAAGAAAG--TCACCTTCAAAAATGCAAAAATCTATAGCTCTGATGTTGTCAG 872
   ||||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 177 LeuLeuGlnProLeuArgGlyLysSerArgAlaSerLeuTrpGlyProAsnIleGlyArg 196
QY 873 CTTGGAAGAAAGCGGCTAAGATGCTGAAGAGCTTCTGGAAGCTGTGGAAGATGAT 932
   ||||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 197 ProArgLysAsnValIleAlaLeuLeuAspGlyPheMetLysValAlaGlySerThrVal 216
QY 933 GATTCAGTTACAGGCGATCTACTATTTGAATGAGAGAGCTGTACAGGAAGATTT 992
   ||||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 217 AspAlaValThrTrpGlnHisCysTrpLysAspLysArgValValLysValMetLysPhe 236
QY 993 CTAAACCCCTGATGATGACATTTTATTTTCACTCTGCAAAAAGTTTCCAGTGCT 1052
   ||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 237 LeuLysThrArgLeuLeuAerPheThrLeuSerAspGlnIleArgLysIleGlnLysValVal 256
QY 1053 GAGAGACACAGGCGCTGCAAGAAAGCTGTGTTAGAGAAAACAAGCTGTGATATGAGGC 1112
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 257 AsnThrLysThrProGlyLysLysIleTrpLeuGlnLysValValThrThrSerAlaGly 276
QY 1113 GAGAGCGCCCTGTATCCGACACCTTTGACGCTGGCTTTATGAGGCTGATAAATGGGC 1172
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 277 GlyThrAsnAsnLeuSerAspSerTrpAlaIleGlyPheLeuTrpLeuAsnThrLeuGly 296
QY 1173 CTGTGACGCCGATAGAGATAGAGATGATGAGGAGCAAGATTTCTTTGAGAGCAAGAAC 1232
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 297 MetLeuAlaAsnGlnGlyIleAspValValIleArgHisSerPhePheAspHisGlyTrp 316

```

```

QY 1233 TACCAATTAGTGAGTAAAGAACTTGATCTTTTACCTGATTAATGAGTATCTTCTGTT 1292
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 317 AsnHisLeuValAspGlnAsnPheAsnProLeuProAspTrpTrpLeuSerLeuLeuTrp 336
QY 1293 AAGAAATTTGGTGGGACCAAGGCTTTAATGCGAAGGTGCGAAGCTTCAAGAAAG--- 1349
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 337 LysArgLeuIleGlyProLysValLeuAlaValHisValAlaGlyLeuGlnArgLysPro 356
QY 1350 -----AAGCTTGCAGTATACCTTCAATTCATTCAGCAAAACACTGAC 1385
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 357 ArgProGlyArgValIleArgAsnLysValLeuArgLysIleTrpAlaHisCysThrAsnHis 376
QY 1386 AATCCAGGATTAAGAGAGAGATTTTACTGTGATGCAATTAACCTCCATTAAGCTGACC 1445
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 377 AsnHisAsnTrpValArgGlySerIleThrLeuPheIleAsnLeuHisArgSerArg 396
QY 1446 AACTACTGCGGCTTACCCCTATCTTTTCTTAACCAAGGTGATTAATGCTTTGATAA 1505
   ||||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 397 LysLysIleLysLeuValaGlyThrLeuArgAspLysLeuValHisGlnTrpLeuLeuGln 416
QY 1506 CCTTTGGGACCTTCATGATTAATCTTCCAAATGCTGCAATGCTCAATGCTTAAG 1565
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 417 ProTrpGlyGlnGlnLysLysSerLysSerValGlnLeuAsnGlyGlnProLeuVal 436
QY 1566 ATGTGAGATGATCAAACTTGGCCACTTAAATGAAGAAACCTCTCCGSCCAGAGATTCA 1625
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 437 MetValAspAspGlyThrLeuProGlyLeuLysProArgProLeuArgAlaArgThr 456
QY 1626 CTGGGCTGCGCAGCTTTCATATCTTTTGTGATTAAGAAATCCCAAGTGTGCT 1685
   ||||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 457 LeuValIleProProValThrMetGlyPhePheValValLysAsnValAsnAlaLeuAla 476
QY 1686 TGC 1688
   |||
Db 477 Cys 477

RESULT 2
T45608
hypothetical protein F13G24.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #extc_change 04-Feb-2000
C:Accession: T45608
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.-J.; Voet, M.; Robben, J.; Volckaert, G.; Bar
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z23009
A:Accession: T45608
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-521 <BEV>
A:Cross-references: EMBL:AL133421
A:Experimental source: cultivar Columbia; BAC clone F13G24
C:Genetics:
A:Map position: 5
A:Introns: 53/3; 66/1; 127/2; 177/1; 256/1; 319/2; 361/2; 394/3
A>Note: F13G24.30

Alignment Scores:
Pred. No.: 1,78e-23 Length: 521
Score: 416.00 Matches: 154
Percent Similarity: 42.05% Conservative: 68
Best Local Similarity: 29.17% Mismatches: 164
Query Match: 13,344 Indels: 122
DB: Gaps: 24

US-10-676-079-3 (1-1721) x T45608 (1-521)
QY 285 CTGGGTTCTCCAAAGCTTGTGATCTTGGCCAGAGCTTGTCTCTGCGTACCTGAGTTT 344
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 55 LeuThrArgProLeuLeuThrLysAlaIleLysAlaPheLysProLeuArgGlyLeuArg 74
QY 345 GGTGGGACCAAGACAGACTCTCAATTTTTCGATTCACCAAGAAAGATCAACTTTGAAGAG 404
   ||||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75 GlyLysSerLeuGlnAspGlnValIleTrpAspValGlyAsnLeuLysThr----- 91

```

```

QY 405 AGAAGTACTGGCAATCTCAAGTCAACGAGATATTTGCAATATGATCCATCCCTCT 464
DB 92 -----ProGlyArg----- 94
QY 465 GATGTGGAGAGAGATTACGGTTGGAAATGGCCCTTACCGAGAGCAATGGTACTCCGAGAA 524
DB 95 -----PropheGlnLysMet----- 99
QY 525 CACTACAGAAAAGTTCAAGACAGC-----ACTTACTCAAGAACCTCTGTA--- 572
DB 100 -----AsnSerGlyLeuPheGlyPheSerLysGlyCysLeuHis 112
QY 573 -----GATGTGCTATACACTTTTGGCAACCTGTCAGAGCTGGAGCTGATTCCTT 620
DB 113 MetLysArgTPrpArgGlnLeuAsnSerPheLeuThrAlaThrGlyAlaValValThrPhe 132
QY 621 GGCCTAAATGCGTTATTAAGAACAGACAGATTGACG-----TGG 659
DB 133 GlyLeuAsnAlaLeuArgGlyArgHisLysLysArgGlyLysAlaArgGlyLysAlaArg 152
QY 660 AACAGTTCTTAATGCTCAGTTCCTCGACCTACTGCTCTTCCAGAGGCTATTAACATT--- 716
DB 153 AspHisLysLeuAsnThrGlnAspPheLeuAsnTyrThrValSerLysGlyTyrValIleAsp 172
QY 717 TCTTGGGAACCTAGGCAATGACCTTAACAGTTTCCTTAAGAGGCTGATATTTTCATCAAT 776
DB 173 SerTPrpGlnPheGlyAsnGlnLeuSerGly-----SerGlyValGlyAlaSerValSer 190
QY 777 GGGTGGCACTTGAAGAGATTATATTCATTGATTAACCTTCTTAAGAAAGTCCACCTTC 836
DB 191 AlaGlnLeuTyrGlyLysAspLeuIleValIleLysAspValIleAsnLys---ValTyr 209
QY 837 AAAAATGCAAAATCTCATGCTCTCATGTTGCTGACCTT-----CGAAG 881
DB 210 LysAsnSerTPrpLeuHisLysArgProIleLeuValAlaProGlyGlyPheTyrGlnGln 229
QY 882 AAGACGGCTGATAGAGCTGAGAGCTTCCGAGAGCTGAGAGATGATGATTCAGTT 941
DB 230 TPrpTyrThrLysLeuLeuGlnIle-----SerGlyProSerValValAspValVal 246
QY 942 ACATGGCATCTACTATATTGAATGACGAGCT-----GCTACCGAGAGATTTTCTA 995
DB 247 ThrHisHisIleTyrAsnLeuGlySerGlyAsnAspProAlaLeuValLysLysIleMet 266
QY 996 AACCTGATGATATTGACATTTTATTTATTCCTGTCGCAAAAAGTTTTC----- 1043
DB 267 AspProSer-----TyrLeuSerGlnValSerLysThrPheLysAspValAsn 282
QY 1044 CAGGTGGTTGAGAGCACAGGCTGGCAGAGAGGTCTGGTTAGGAGAAACAGCTCTGCA 1103
DB 283 GlnThrIleGlnGlnHisGlyProTPrpAlaSerProTPrpValGlyGlnSerGlyAla 302
QY 1104 TATGAGGCGGAGCGCCCTGCTATCCGACCTTTCAGCTGAGCTTATATGCGCTGAT 1163
DB 303 TyrAsnSerGlyGlyArgHisValSerAspThrPheIleAspSerPheTPrpLeuAsp 322
QY 1164 AAATGGGCTCTGACCCCGAATGGGAATGAAGTGTATGAGGAGAGTATTTCTTGA 1223
DB 323 GlnLeuGlyMetSerAlaArgHisAsnThrLysValTyrCysArgGlnThrLeuValGly 342
QY 1224 GGAGGAAATACCATTTATGATGAGTAA---AACTTGATCTCTTATCTGATTTAGGCTA 1280
DB 343 ---GlyPheTyrGlyLeuLeuGlnLysGlyThrPheValProAsnProAspTyrTyrSer 361
QY 1281 TCTCTTCTTCAAGAAATGTTGGTGGCAGACAGTGTATATGGCAGCGTGCAGAGTTCA 1340
DB 362 AlaLeuLeuTPrpHisArgLeuMetGlyLysGlyValIleValAlaGlnThrAspGlyPro 381
QY 1341 AAGAGAGAGAGCTTGAATATACCTTCAATTGACAAAACATGACATCAAGTATTA 1400
DB 382 Pro-----GlnLeuArgValTyrAlaHisCysSerLys-----Gly 393

```

```

QY 1401 GAAGGAGATTAACTCTGTATGATCCATAAACCCTCCATTAACGTCCACCAAGTACTTG----- 1454
DB 394 ArgAlaGlyValThrLeuLeuLeuIleAsnLeuSerAsnGlnSerAspPheThrValSer 413
QY 1454 ----- 1454
DB 414 ValSerAsnGlyIleAsnValValLeuAsnAlaGlnSerArgLysLysSerLeuLeu 433
QY 1455 ---CGGTACCCCTATTCCTTTTCT-----AACAGCAAGAGGATTAATACCTCTTA 1502
DB 434 AspThrLeuLysArgProPheSerTPrpIleGlySerLysAlaSerAspGlyTyrLeuAsn 453
QY 1503 AGACCT-----TTGGACCTCATGCA-----TTACTTTCCAATCTGTCCACTC 1547
DB 454 ArgGlnGlnTyrHisLeuThrProGlnAsnGlyValLeuArgSerLysThrMetValLeu 473
QY 1548 AATGCTTAATCTTAAGAAGTGTGATGATCAAACTTGGCCACTTATATGAAAAAACC 1607
DB 474 AsnGlyLysSerLeuLysProThrAlaThrGlyAspIleProSerLeu---GlnProVal 492
QY 1608 CTCGGGACCA---GCAAGTCACTGGGCTTGGCAGCTTCTCATATATGTTTTTGTGATA 1664
DB 493 LeuArgSerValAsnSerProLeuAsnValLeuProLeuSerMetSerPheIleValLeu 512
QY 1665 AGAAATGCCAAAGTTGCTGCTTGC 1688
DB 513 ProAsnPheAspAlaSerAlaCys 520

```

RESULT 3

T01953
 hypothetical protein T2L5.6 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 21-Jan-2000

R:Geisel, C.; Smith, A.; Le, T.
 submitted to the EMBL Data Library, October 1998
 A:Description: The sequence of A. thaliana T2L5.
 A:Reference number: Z14470

A:Accession: T01953

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-190 <GB>

A:Cross-references: EMBL:AF096371; NID:g3695386; PID:g3695392

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 4

A:introns: 36/2; 69/3

A>Note: T2L5.6

C:Superfamily: Arabidopsis thaliana hypothetical protein T2L5.6

Alignment Scores:
 Pred. No.: 6,86e-05 Length: 190
 Score: 169.50 Matches: 54
 Percent Similarity: 45.36% Conservative: 34
 Best Local Similarity: 27.84% Mismatches: 57
 Query Match: 5.43% Indels: 49
 DB: 2 Gaps: 9

US-10-676-079-3 (1-1721) x T01953 (1-190)

```

QY 1206 AGCAAGATATTTCTTGGAGCAGAGAACTACCATTTAGTGAAT---GAATACTTCATCT 1262
DB 12 ArgGlnSerLeuIleGly---GlyAsnTyrGlyLeuLeuAsnThrThrAsnPheThrPro 30
QY 1263 TTACTGTATATTTGGCTATCTCTTCTGTTCAAGAAATTTGGTGGCAGCAGGTGTTATG 1322
DB 31 AsnProAspTyrTyrSerAlaLeuIleTPrpArgGlnLeuMetGlyArgLysAlaLeuPhe 50
QY 1323 GCAAGCGTGCAGAGTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1382
DB 51 ThrThrPheSerGlyThrLys-----LysIleAsnSerTyrThrHisCysAla----- 66
QY 1383 GACAAATCCAGAGTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1442

```

```

Db      67 -----ArgInserTgLyIleThrValLeuMetAsnLeuAspAsnThr 82
Qy      1443 -----ACCAAGTAC 1451
Db      83 ThrThrValValAlaValValGluLeuAsnAsnSerPheSerLeuAlaGlnIleValHis 102
Qy      1452 TTGGGGTACCCCTTAATCTTTTCTTACAGCAAGCATGTAAATPACTCTTAAGACCTTTG 1511
Db      103 MetLys-----SerTyrIleArgIleSerSerGlnLeuPheGly----- 115
Qy      1512 GAGCCCTATGATTAATCTT----- 1529
Db      116 GlyProAsnGlyValIleGlnArgGluGlnThrIleLeuThrAlaValAspGlyAsnLeu 135
Qy      1530 ---TCCAAATCTGTCCAACTCAATGATCTTAAGTAAAGATGGTATGATCAAACTTG 1586
Db      136 HisSerGlnThrMetLeuAsnGlyAsnAlaLeuGlnValAsnSerMetGlyAspLeu 155
Qy      1587 CCACTTTAATGAGAAACCTTCGGCCAGGAATTCAGTGGGCTTGCAGCTTTCTCA 1646
Db      156 ProProIleGluProIleHisIleAsnSerThrGluProIleThrIleAlaProTyrSer 175
Qy      1647 TAAAGTTTTTTGGAGAAAGAAAGCCAAAGTTCCTGCTTGC 1688
Db      176 IleValPheValHisMetArgAsnValValValProIaCys 189

```

RESULT 4

S74760 hypothetical protein slr1617 - Synechocystis sp. (strain PCC 6803)

C/Species: Synechocystis sp.

A/Variety: PCC 6803 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C/Accession: S74760

R/Kanehara, T.; Sato, S.; Korani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

8.

A/Reference number: S74322; MUID:97061201; PMID:8905231

A/Accession: S74760

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-411 <KAN>

A/Cross-references: EMBL:D90901; GB:AB001339; NID:G1651897; PIDN:BA16911.1; PID:dl01764

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Alignment Scores:

Pred. No.:	0.035	Length:	411
Score:	135.00	Matches:	89
Percent Similarity:	33.46%	Conservative:	72
Best Local Similarity:	21.81%	Mismatches:	122
Query Match:	4.33%	Indels:	125
DB:	2	Gaps:	22

US-10-676-079-3 (1-1721) x S74760 (1-411)

```

Qy      287 GGGTTTCAAGAGCTTGTGTCAGAGAGCTTGTCTGCTGCTGACCTGAGTTGG 346
Db      67 GlyMetSerArgSerAsnAlaIleAlaValGlyIleThrAsnLeuGluValIlePro 86
Qy      347 TGGCACCAG-----GAC 358
Db      87 PheAsnGlnAlaLeuArgGlnValGlnGlnAlaThrGlyGlnAlaGlnIleLysLeuAsp 106
Qy      359 AGACTTCCTAATTTTGTGATCCCAAGAGATCACTTTGAGAGAGATTACTGCA 418
Db      107 ProLeuAsnAlaValAlaGlnAlaLeuGlyValAlaAsnIleLysGlyLys----- 123
Qy      419 ATCTCAATGCAACAGATATTGTCAAATATGATTCATCCCTTCCTGATGTGAGAGAA 478
Db      124 ---CysSerGln-----TyrArgAsnIleLeuIle-----ThrGlyGlyArg 136

```

```

Qy      479 GTTACGGTTGAATGCGCTTACCGAGGCAATTTGCTACCGAGAACTATCCAGAAAA 538
Db      137 GlyPheIleGlyThrAlaLeuGlnGlnIleAla-LeuIle-----AsnSerG 151
Qy      539 GTTCAAGAACAGACCTTCAAGAGCTCTGAGATGTCATACCTTTTGCAACAG 598
Db      151 uPheArgLeuIleSerProThrArgGlnIleAlaPheIle-----PheAlaLys 168
Qy      599 CTCAGAGCTGACCTTATCTTT-----GGCTTAATGCGTTATTAAGAACAGACAG 649
Db      168 rThrIlyLeuAspLeuLeuAlaSerGluGluAsnIleAspCysIleValHisLeuAla 188
Qy      650 TTGCACTGAGAACAGTTCTAATCTCACTTG-----CT 682
Db      188 nProArgValIlyrThrSerAsnValAlaMetGlyGlnThrIleuThrMetLeuArgAsnVal 208
Qy      683 CCGGACTACTGCTCTTCCAGAGG-----TATTAATTTCTTTGGAACTA--- 728
Db      208 ILeAspValCysLeuAlaLysAspIleProLeuIleTyrProSerSerTyrGlnIleTyr 228
Qy      729 -----GGCAATGAACTTAACAGTTTCCTTAAGAG----- 758
Db      228 rSerGlyTyrAlaGlyThrIleHisAlaAspGluSerThrProAlaLeuProArgIlyr 248
Qy      759 -----CTGATATTTTCATCAATGGGTGCGCACTTAAGAGAGA 796
Db      248 oTyrGlyGluThrIlyrTyrLeuAlaGlnIleLeuIle-----As 261
Qy      797 TTATATTCATTCATCAATCTTCAAGAAAGTCCACTTAAGAAATGCAAACTATAG 856
Db      261 rHisCysArgArgThrArgGlyLeuArgCysAlaAlaLeuAspSerSerProValIlyr 281
Qy      857 TCCGTATGTTGTGACCTTCGAGAAAGACGCGTAAGATGCTGAGAGCTTCGTAAG 916
Db      281 ySerMetSerAspIlyrPro-----LysPheIlePheAsnPheLeuIlyr 296
Qy      917 TGTGAGAAAGATGATGATTCAGTTACATGCGCTCACTATTTTAAAGAGAGCTGC 976
Db      296 sAlaSerGlnGlyGlnIlyrIleValIlyr-----HisIlyrIleAsnGly----- 311
Qy      977 TACCAGGAAGATTTCTAAGCCCTGATGA-----TTGCAATTTTATTTTC 1024
Db      312 -----AsnProLysIleuAspLeuHisIleAspAspLeuIleSe 325
Qy      1025 ATCTGTGCAAAAAGTTTCCAGGTGTTGATGAGACACAGCCTGCGAGAGAGCTGCT 1084
Db      325 rSerIleValAlaThrLeu-----LysSerArgPheI 336
Qy      1085 AGAGAAACAGCTCTGCATATGAGAGCGGAGCGCCCTTGTATCCGACCTTTGCGAC 1144
Db      336 eGlyAsnLeuAsnIle-----GlyThrGlyGlnIleuSerSerThrIleuLysI 352
Qy      1145 TGGCTTATGATGCGTGAATAATGGCGCTGCGAGCC-----CGAATGGG 1189
Db      352 eAlaGluMetIleArgAspGluLeuGlySerSerSerMetIleGlnGlnIleGluValAs 372
Qy      1190 AATGAAGTGTGATGAGCAAGATATCTTTGAGAGAGAAATACATTAATGATGATGA 1249
Db      372 nThrGluValAlaSerIleAlaMetAsnTyrGlyArgAlaAsn-----HisValLeuAsp- 390
Qy      1250 AAACCTTCATCTTACCTGATATTTGCTATCTCTTCCTGTTCAAGAAATGGTGGGCA 1309
Db      391 -----TyrIleProValIlePhePheGlyu-----GlnIlyr 401
Qy      1310 CAAGTGTATATGCGAAGCGTG 1331
Db      401 uLysSerLeuLeuHisGlnIle 408

```

RESULT 5

T10050 integrin alpha-v chain precursor - mouse
N/Alternate names: vitronectin receptor alpha chain
C/Species: Mus musculus (house mouse)

CjDate: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 CjAccession: T10050
 RjMada, J.; Kumer, A.; Liu, Z.; Ruoslahti, E.; Reichardt, L.; Marvaldi, J.; Kanwar, Y.S.
 J. Cell Biol. 132, 1161-1176, 1996
 A>Title: Cloning of mouse integrin alpha v cDNA and role of the alpha v-related matrix r
 A;Reference number: Z16920; MIMD:16176109; PMID:8601592
 A;Accession: T10050
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-1044 <MAD>
 A;Cross-references: EMBL:U14135; NID:G537490; PIRID:AA052497.1; PID:G537491
 A;Experimental source: strain CD-1, kidney
 C;Superfamily: integrin alpha-2b chain
 C;Keywords: cell adhesion; receptor

Alignment Scores:

Prod. No.:	0.281	Length:	1044
Score:	124.50	Matches:	137
Percent Similarity:	35.97%	Conservative:	86
Best Local Similarity:	22.10%	Mismatches:	213
Query Match:	3.99%	Indels:	184
DB:	2	Gaps:	35

US-10-676-079-3 (1-1721) x T10050 (1-1044)

```

QY 48 CCAAGTGAACCAAGTGTGCTGCGCTGCAAGCTGCGCGCGCGCTGATGCTG 107
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 4 ProGly-----ArgLeuLeuLeuArgProArgProGly-----GlyLeuLeuLeu 18

QY 108 CTGCTCTGGGGCGCGCTGGGCTCTCTCCCTGGCGCGCTGCGCGCTGCGAAGCA 167
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 19 LeuLeuProGlyLeuLeuLeuProLeuLeuA----- 28

QY 168 CAGGAGCTGTGAGCCTGGAATCTTTCACCCAGAGCGCGCTGCTGTGAGCCCC 224
    |||||  ||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 29 --AapAlaPheAsnLeuAsp-----ValGluSerProAlaGluTyrAlaGlyProGlu 45

QY 225 ---TCGTCCTGCTCCGTCACATTTGAC---GCCAAGCTGACCAAGCGCGCGCTCTC 278
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 46 GlySerTyrPheGlyPheAlaValAspPhePheGluProSerThrSerSerArgMetPhe 65

QY 279 ATCTCTCGGGTCTTCCAAAGCTTCTGACCTTGCGCGAAGCTTGTCTCTGCGTAC 338
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 66 LeuLeuValGlyAlaProGlyAlaLeuThrGlnProGlyAlaGlu----- 82

QY 339 AGGTTGGTGGCAGCAAGCAAGCTTCTTAATTTTGATCCCAAGAAATCAACCTTT 398
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 83 -----GlyGlyGlnValLeuLeuValCysGlyCys--SerSerSerArgArgCysGlnPro 100

QY 399 GAAGAGAAAGTTAAGTCAATCTCAAGTCAAGCAAGTATTTGCAATATGATTCATC 458
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 100 eGluPheAspSerThrGlyAsn-----ArgAspTyrAlaValAspAspPro-- 115

QY 459 CCTCTGATGTGAGAGAGTTCACGTTGAGATGAGCCCTTACAGAGCAATTTGCTACTC 518
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 116 -----LeuGluPheLeuSerHisGlnThrPheGlyAla----- 126

QY 519 CAGAGAACCTACCAAAAAAGTTCAAGAACAGCACTTCAAGAAAGCTCTGATGTCG 578
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 127 -----SerValArgSerLys--GlnAsp-----LysValLeuAlaCysAs 139

QY 579 CT-----ATACACTTTTGCAAACTGCTCAGAGCTGAGCTGATCTTTGCGCTAAATGCG 632
    |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 139 LeProLeuTyrHisIlePheArgThrGluMetLeuGlnGluArgGluProValGlyThrCysAs 159

QY 633 TTATTAAAGAACAGCATTTGACGTGAACAGTTCTATGTC-----T 674
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 159 heLeuGlnHisPheGlyThrLysThrValGluTyrAlaProCysArgSerLysAsnIleAspAs 179

QY 675 CAGTGTCTCTGAGCTACTGCTCTTCCAGAGGTTAAACATTTCTTGGGAACCTAGCAAT 734
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 179 IaAspGlyGlnGly--PheCysGlnGly--GlyPheSerIleAspPhe----- 193
  
```

```

QY 735 GAACCTTAACAGTTTCCTTAAGAGCTGATATTTTCATCAATGGGTCCGACTTGAAGAA 794
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 194 -----ThrValAlaAspArgValLeuGlyGlyPro---GlySer 206

QY 795 GATTATTAATTCATTCATTAATTTCTTAAGAAAGTCCACTTCAAAAATGCAAACTTAT 854
    |||||  ||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 207 PheTyrTrpGln--GlyGlnLeuLeuSerAspGlnValAlaGlnIleIleSerLysTyr 225

QY 855 GGTCCTGATGTT-----GGTCAAGCTCGAAGAAAGCGGCTAAGAG 896
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 226 AspProAsnValTyrSerIleLysTyrAsnGlnLeuAlaThrArgThrAlaGlnAla 245

QY 897 CTGAAGACTTCTGAAAGCTGTGAGAGTGAATGATTCAGTTACATGCGATCCTAC 956
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 246 Ile-----PheAspAspSer 250

QY 957 TATTGAAATGACGAGACTGTAACAGGAAAGATTTTCTTAACCTGATGTATGACATTT 1016
    |||||  ||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 251 TyrLeu--GlyTyrSerValAlaValGlyAspPhe--AsnGlyAspGlyIleGluAsp 268

QY 1017 TTTATTGATCTGTCGCAAAAAGTTTTCAGAGTGTGAGAGAC-----AGGCTGGC 1070
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 269 PheValSerGlyValProArgAlaAlaArgThrLeuGluMetValTyrIleTyrAspGly 288

QY 1071 AAGAAGTTC-----TGCTTGAAGAGAAACAAGCTCTGATATGAGAGC--- 1112
    ||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 289 LysAsnMetSerSerLeuHisAsnPheThrGlyGlnGlnMetAlaIleTyrPheGly 308

QY 1113 -----GAGCGGCC 1121
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 309 SerValAlaAlaThrAspIleAsnGlyAspAspTyrAlaAspValPheIleGlyAlaPro 328

QY 1122 TTGCTATCCGACACTTTGACAGCTGGC---TTTATGTCGGGATTAATTTGGCGCTGCA 1178
    |||||  ||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 329 LeuPheMetAspArgGlySerAspGlyLysLeuGlnGluValGlyGlnValSerValSer 348

QY 1179 GCCCGAATG-----GGAATGAAGTGTGATG 1205
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 349 LeuGlnArgAlaValGlyAspPheGlnThrThrLysLeuAsnGlyPheGluValPheAla 368

QY 1206 AGGCAAGTATTTCTTGAAGCAAGAAATCAATTTAGTGAATGAATCTTGATCTTTTA 1265
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 369 Arg-----PheGlySerAlaIleAlaProLeuGlyAspLeuAspGlnAspIlyPhe 385

QY 1266 CCTGATTAATGGCTATCTCTTCTGTCAAGAAATTTGGGCGCAAGAGTGTAAATGCGCA 1325
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 386 AsnAsp-----IleAlaIleAla 391

QY 1326 AGCGTCAAGGTTCAAGAGAAAGAGCTTGAATATCTTCAT----- 1370
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 392 AlaProTyrGlyGlyGlnAspIlyLysGlyLeuValTyrIlePheAsnGlyArgSerThr 411

QY 1371 TGCACAAACACTGACAAATCAAGTATTAAGAGAGATTTAACTCTG----- 1418
    |||||  ||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 412 GlyLeuAsnSerValProSerGlnIleLeuGlnGlyGlnTyrAlaAlaGlnSerMetPro 431

QY 1419 -----TATGCATTAACCTCATTAAGCTCAAGTACTTGGCTTACCTATCT--- 1469
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 432 ProSerPheGlyTyrSerMetLysGlyAlaThrAspValAspAspAsnGlyTyrProAsp 451

QY 1470 -----TTTTCTTAACAGAGTGAATTAATCTTCTTAAGACTTTTGGA 1514
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 452 LeuValValGlyAlaPheGlyValAspArgAlaValLeuTyrArgAlaArgProVal--- 470

QY 1515 CCTCATGATTAATTTCAAATCTGTCAACTCAAT---GGTCTTAAGCTTAAG----- 1565
    |||||  ||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 471 -----ValThrValAsnAlaGlyLeuGlnValTyrProSer 482

QY 1566 ---ATGCTGATGATCAACCTTGGCACTTTAATGAAAAAAGCTCTCGCGGAGAGT 1622
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 483 IleLeuAsnGlnAspAsnLysIleCysProLeu-----ProGlyThr 496

QY 1623 TCACGTGGGCTTGCAGCTTTCATATAGTTTGTGATTAAGAAATGCCAA 1676
  
```

Db 497 AlaLeuValSerCysPheAsnValArgPheCysLeuValAspGlyLys 514
 RESULT 6
 T49648
 hypothetical protein B8B20.20 [imported] - Neurospora crassa
 C/Species: Neurospora crassa
 C/Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
 C/Accession: T49648
 R/Schulte, U.; Algen, V.; Hohenseel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
 submitted to the Protein Sequence Database, May 2000
 A/Reference number: Z25022
 A/Accession: T49648
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-2298 <SCH>
 A/Cross-references: EMBL:ALJ55933; GSPDB:GN00116; NCSP:B8B20.20
 A/Experimental source: BAC clone B8B20; strain OR74A
 C/Geneticist:
 A/Map position: 6
 A/Introns: 426/3

Alignment Scores:
 pred. No.: 0.99 Length: 2298
 Score: 118.50 Matches: 117
 Percent Similarity: 32.99% Conservative: 73
 Best Local Similarity: 20.31% Mismatches: 185
 Query Match: 3.80% Indels: 201
 Gaps: 27
 DB: 2

US-10-676-079-3 (1-1721) x T49648 (1-2298)

QY 316 GAGGCTGTCTCTCGCTACCTGAGTGTGGTGCCACCAAGACAGCTTCTAATTTTCG 375
 Db 1464 GluAlaCysLeuIleAsnIle-ArgAlaTrpAsnGlnLeuAlaArgLeu-ValValSerA 1483
 QY 376 ATCCCAAGAAAGAACTCACTTGAAGAGAAAGTTACGCAATCTCAAGTC--AAC 432
 Db 1483 snlglgluglyserAlaSerPheArgPheProPheIleThrTrpArgAsnAsnValPheAsnG 1503
 QY 433 AGGATATTGCAATATGATTCATCCCTCTGATGAGGAGAAATTTCG-- 485
 Db 1503 InIleuAsnArgInIlyMetSerAlaGlnSerAspIleGlnGlnIlePheArgAlaLeuS 1523
 QY 486 -----TTGGAAATGSCCTACGACGAGCAATTCCTCCGAGAAC 525
 Db 1523 eraIaGluAsnMetArgSerIleAspAlaIaIaTrpArgGlnGlnIleThrIlyAsn- 1542
 QY 526 ACTAACCAAAAAAGTTCAAGAACGACCTACAGAGCTCTGATGATGTCTATACA 585
 Db 1543 -----LysAlaThrAlaLeuAspIleLeuHisIer 1552
 QY 586 CTTTGGCAACTGCTCAGACTGAGCTGATCTTTGGCCTAAATGCGCTA-- 635
 Db 1552 hrSerAlaArgAlaSer--LeuAspValIleuLysGlnAlaIyThrIleuGlnAlaIaI 1571
 QY 636 -----TTAAGAACGACAGATTTCAG-- 656
 Db 1571 LeIyThrIleuAsnValThrGlnLeuGlnLysMetCysThrThrIleuHisPheGlySerP 1591
 QY 657 -----TGGACAGTTCTTAATGCTCAGTTGCTCTCGACTACTGCTCTTCCAAAG 705
 Db 1591 roGlyPheAspTrpGlyIleLeuAsnValAlaLeu-----AspThrHisA 1606
 QY 706 GGTATAACATTTCTGTG--GACTAGCGAATGAACTAAACAGTTTCTTAAGAGGCTG 762
 Db 1606 laHisPheLeuGlyTrpIleGlnThrSerSerGlnGlnIlyrSerSerAsnGlnSerS 1626
 QY 763 ATATTTTCATCAATGGGTCCAGTTAGAGAAAGATTATTCATATTCATTAACCTTTAA 822
 Db 1626 eraIaAspIleAspProArgGlnLeuGlnuAspAlaIleLeuLeuLeuGlnIlyuLysLeuT 1646

QY 823 GAAAGTCCACCTTGCAAAAAATGCAAAACCTATATGCTCTGATGTGTGCTACCTCGAAGA 882
 Db 1646 hrIySgIlyPhePheTrpMetAla-----ArgGluL 1656
 QY 883 AGACGGCTAAAGATGCTGAAGAGCTTCCTGAAGCGTGGAGAGAGATTGAT--TCAG 939
 Db 1656 euLeuAlaLeuProLeuLysAlaIleThrThrPheGlySgIlyGlnIlyGlnValAlaC 1676
 QY 940 TTACATGAGCATCTACTATTGAAATGACGGACTGCTACAGGAAAGATTTTCTAACC 999
 Db 1676 yethrGlnLysThrValThrIleuAlaIaIyLeuAlaIaArg----- 1690
 QY 1000 CTGATGATATGACATTTTATTTATCTCTGTGCAAAAAGTTTCCAGTGCTTGAGAGCA 1059
 Db 1691 -----PheIleGln-----GluArgValThrGlnValIleuProTyP 1703
 QY 1060 CCAGCGCTGAGCAAG-----A 1074
 Db 1703 heGlnProGlyLysTrpGlyLeuPheProAspMetProLysAsnMetSerGlyProGluA 1723
 QY 1075 AGCTGTGTTAGAGAGAAACAGCTCTGATATGAGCGCGAGCGCTTGTATCCGACA 1134
 Db 1723 rgrArgTrpLeu-----ProLeuPheIleAlaI 1732
 QY 1135 CCTTTCAGCTGCGC-----TTTATGT 1155
 Db 1732 hrLeuValAsnLysAsnValPheAspPheLysAspIleGlnThrAsnIleLeuSerLeuT 1752
 QY 1156 GGCTGATTAATGATGGCGCTGTGACCGCA--ATGGGAATAGAAAGTGATGAGGCAAG 1212
 Db 1752 rpaIaGlnSerIleIleLysProMetArgPheLeuGlyTrpGlnThrTrpLeuAlaGlnI 1772
 QY 1213 TATTTCTTGGAGCGAAGAACTACCATTTA-----GTGATGAAATCTTGATCCTTTAC 1266
 Db 1772 alLeuGlnGlnArgLysLeuProPheLeuAlaGlnIlaAspValSerAlaGlyMetThrP 1792
 QY 1267 CTGATTAATGGCTATCTCTT--CTGTTCAGAAATATGTTGGGC-----ACCAAGTGT 1317
 Db 1792 roAspTrpAsnIleHisLeuAspLeuPheSerArgAlaIleHisTrpMetArgLysAlaI 1812
 QY 1318 TA-----ATGGCAAGCGTGCAAGTT 1338
 Db 1812 euArgGlyLysAlaThrThrProAlaProGlyValThrSerAlaSerThrAlaGlyS 1832
 QY 1339 CAAG-----AGAGGAGCTTCGAGTATACCTTCATTCGACAAACATGT 1383
 Db 1832 erSerSerAlaGlnSerIleArgArgGlnArgGlnIlyPheSerHis----- 1847
 QY 1384 ACAATCCAAAGTTTAAAGAGAGATTTAATCTGTATGCCATTAACCTTCATACGTCA 1443
 Db 1847 ----- 1847
 QY 1444 CCAAGTATCTGGGGTATACCTATCTTTTCTAACAAGCAAGTGATAATACCTTCTAA 1503
 Db 1848 -----ThrLeuGlnLeuAlaMetThrAsnIleLysLysAspLeuPheLeuLeu 1864
 QY 1504 GACCTTTGGGA----- 1514
 Db 1864 rgsrLeuAlaLeuAlaAspProThrAlaSerSerThrGlnIlyHisArgAspTyPMetA 1884
 QY 1515 -----CCTCATGATTAATCTTCCAAATCTGTCCAACTCAATGCTTAACATGAG 1569
 Db 1884 laPheThrHisGlyLeuIleSer--LeuIleLysSerHisGlyValGlyIleValVal 1903
 QY 1570 TGGAT-----GATCAACCTTGCCACCTTAAATGAGAAAACCTG 1608
 Db 1903 alAspSerPhePheLeuThrProSerAspSerTySerProLeuGlnAspProGlnH 1923
 QY 1609 TCCGGCCAGGA-----AGTTCACTGGGCTTCCAGCTT 1641
 Db 1923 euHisThrAlaGlyIleMetAlaTyGlyValArgLeuSerGlnLysAspValProAlaA 1943
 QY 1642 TCTCATATAGTTT---TTTGATATAAGAAATGCCAAAGTTGCT 1682

Score: 112.50 Matches: 85
Percent Similarity: 33.17% Conservative: 48
Best Local Similarity: 21.20% Mismatches: 143
Query Match: 3.61% Indels: 125
DB: 2 Gaps: 18

US-10-676-079-3 (1-1721) x F64383 (1-356)

QY 438 ATTGCAAAATAT-----GATCCATCCCTCTGATGAGAGAAATTA 482
Db 14 IleaArgysrYrmetLysIleTyraSngIyLysaSnGIuLysaSpIleYsgIuArgLeu 33
QY 483 CGGTGGAATGCCCCCTACAGAGCAATTGCTACTCCGAGAACCTACCAAGAAAATTC 542
Db 34 ILeysgIu-----LeuYsgIuGluHISValLeuValGluThr 46
QY 543 AAGACAGCAGCTTACTCAGAAAGCTCTAGAT----- 575
Db 47 GluHspolYrHrYrHrLeuLysaIleGluHspolGluHwectMetHsSerLysVal 66
QY 576 -----GTCTATACACTTTTGCAAACTGCTCAGAGCTGAGCTTGATCTT 620
Db 67 GlYAlaLeuYsgIuAlaIleTyLysPheHlaLysProSer----- 80
QY 621 GGCCTAAATGCGTTATTAAGAACGACGATTGCAAGTGAACAGCTTAAATGCTCACTG 680
Db 81 -----LysIleThraSpIeu-----SerAsnProArg-- 89
QY 681 CTCCTGAGTACTGCTCTCCAGAGGATTAACATTTCTTGCAACTGAGCAATGAACT 740
Db 90 ValLeuHspIeuCySerGlyMetGlyYrHnHAlaIleAlaIleHISrYrAsnLys 109
QY 741 AACAGT-----TTCTT 752
Db 110 AsnAlaGluIleAspMetValGluIleCySgIuGluValLeuPheLeuThrLeuPheLeu 129
QY 753 -----AAGAGGCTGATATTTTCATCAAT 776
Db 130 AspIleProTyryLysGluHISgIuIleIleLysaSpIyValArgGluTyRPhleLeuAsn 149
QY 777 GGGTCGAGTGAAGAGAAATATATTCATTCATTAACCTTAAAGAGTCCACCTTC 836
Db 150 -----LysIleGlyIleGluTyR-----LysSerAspTyR 159
QY 837 AAAAATGCAAAATCTATGCTGATGTTGTCAGCTCGAAGAAAGACGCTAATG 896
Db 160 AspAsnIleAsnLeuTyR-----ValGlyAspAlaArgLysPheIleIleLysSer 176
QY 897 CTGAAGAGCTTCCGAAAGCTGTGAGAAAGTATTCATTCATTCATGATGATCACTAC 956
Db 177 AspLysLysTyR-----AsnValIValIlePheHISAspAla 187
QY 957 TATTTGAATGAGACGAGCTGCTACC-----AGGAGATTTTAAACCTGATGATG 1010
Db 188 PheSerProLysaSpIeuProHrLeuTyRThrYrAspPheLeu----- 202
QY 1011 GACATTTTATTCATCTGTGCAAAAAGTTTCCAGGTGTGAGACACAGGCTGCG 1070
Db 203 -----LysGluIleTyRAspMetGluAspAsn-----Gly 213
QY 1071 AAGAGGCTGTGTTAGAGAAACAAGCTGCAATGAGAGCGAGCGCTTCTATCC 1130
Db 214 ValLeuLeu-----SerYrSerAlaIleIleProPheAspSer 226
QY 1131 GACACCTTTCAGCTGCTTATGCTGATGATTAATGCGCTGTGAGCCGAGATGGA 1190
Db 227 AlaLeuValAspCySgIyPheValIleSerGluYsgIuSerValGlyArgLysArgGly 246
QY 1191 ATGAAAGTGTGATGAGCAAGTATCTTTGAGACAGAAACTGATTAATGATGATGAA 1250
Db 247 ILeThrLeuAlaTyryLysaSnProAsnPhelysProAsnArgIleAsnGluValAspGlu 266
QY 1251 AACTTGAT-----CCTTACCTGATTAATGCTATCTCTTCTG 1289

Db 267 ArgValIleAlaLeuSerValIleAlaLeuProTyraArgAspGluThrLeuSerLeuThr 286
QY 1290 TTCAGAAATATGTTGGGACACCAAGGTGTTAATGCAAGGTCGACAGTTCAAAGAAAG 1349
Db 287 LysAspLysIleIleGluAspArgGluHISgIuHISgIuLysLeuLysGluLysLeuIle 306
QY 1350 AACCTTCAGATATACCTTATTCAGCAAACTGACCAATTCAGTATTAAGAGAGAGAT 1409
Db 307 LysIleGlyLysTyRLeuSerThrLysGluIleLysGlyAsnIleProGluGluIle 326
QY 1410 TTAACCTGTATGCC-----ATAACCTTCATACGTACCAAGTACTTGCGTTACCC 1463
Db 327 LeuLysIleGluLysGluAspLeuAsnSerGluIleIleLysLysMetArgLeuLys 346
QY 1464 TAT 1466
Db 347 Phe 347

RESULT 10
S32961
hypoetical protein YBR259 - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypoetical protein YBR1727
C:Species: Saccharomyces cerevisiae
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 19-Apr-2002
C/Accession: S32961; S46140
R:Doignon, F.; Biteau, N.; Crouzet, M.; Aigle, M.
Yeast 9, 189-199, 1993
A:Title: The complete sequence of a 19,482 bp segment located on the right arm of chromosome number: S29348; PMID:93220397; PMID:8465606
A:Reference number: S29348; PMID:93220397; PMID:8465606
A:Accession: S32961
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-688 <DOI>
A:Cross-references: EMBL:X70529; NID:g1907246; PIRN:CAA49923.1; PID:g296558
R:Aigle, M.; Bachel, M.C.; Barthe, C.; Biteau, N.; Crouzet, M.; Doignon, F.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45940
A:Accession: S46140
A:Molecule type: DNA
A:Residues: 1-688 <AIG>
A:Cross-references: EMBL:Z36128; NID:g536684; PIRN:CAA85222.1; PID:g536685; MIPS:YBR259W
C:Genetic8
A:Cross-references: SGD:S0000463
A:Map position: 2R
C:Superfamily: Saccharomyces cerevisiae hypoetical protein YBR259W

Alignment Scores:
Pred. No.: 2.23 Length: 688
Score: 112.00 Matches: 101
Percent Similarity: 35.10% Conservative: 78
Best Local Similarity: 19.80% Mismatches: 168
Query Match: 3.59% Indels: 163
DB: 2 Gaps: 25

US-10-676-079-3 (1-1721) x S32961 (1-688)

QY 438 ATTGCAAAATGATGATCCCTCTGATGAGAGAAATGACGTTGAGATGAGCC 497
Db 164 MetAlaGluTyRserSerTrpLysAspAspLysArgGluLeuGluPheMet 183
QY 498 TACAGAGCAATGCTACTCCGAGAACACTACCAAGAAAGTTCAAGAACAGCACCTTAC 557
Db 184 TyRGlupheArgMetLysLysGluCySLeuValLysPheTyRGlupheAspLeu 203
QY 558 TCAAGAGCT-----CTAATGCTATCACT 587
Db 204 GluLysSerSerAspProLeuLysGluLeuIleIleProTrpGluLysIleValIyVal 223
QY 588 TTGCAAACTGCTCAGACTGACCTTATGCTTAAATGCTTATTAAGAAAGCA 647
Db 224 ---AlaAsnCyS-----IleAspAlaPheMetGlyGluGluValArgIleAspGlyAla 240

```

QY 648 GATTGCAAGTGAACGTTCTAAT-----GCTGAGTGTCTC 683
DB 241 GtuleuiletrThrSerIysasnleuValPheSerSerileSerSerleValleuArg 260
QY 684 CTGGAC-----TACTGCTCTCCAAAGGGGTAT----- 710
DB 261 leuAsnleuGlnAsnMetPheSerAlaPheArgProtyrGlyGlnGluAlaVal 280
QY 711 -----AACATTTCTGGGAAGTGAACGAAAGAACCTTAACGATTTC 749
DB 281 GlnAspPheAlaHisIleArgSerleuIleuArgSerleuIleuValGlnSerleu 300
QY 750 CTTAAGAGGCT-----GATATTTTC-----ATCAATGGCTCGCAGTTAGAA-- 791
DB 301 IleArgAlaIleuIlePheAsnMetPheProtyrPheAsnIleGlnIleValAspThr 320
QY 792 ---GAAGTATATTCATTCGATTAACCTTCAAGAAAG----- 827
DB 321 LysAlaAspGlyIlePhePheleuArgleuLeuArgLysAsnPhelGlnHisIleAsn 340
QY 828 -----TCCACCTTCAAAAT----- 842
DB 341 AspValIleAspPheHisIleGlnValIleLysIleuAsnSerGlnPheLysAsnAsn 360
QY 843 ---GCMAAAGCTTATGCTCTGATGTGTGCTGAGCTCGAAGAAAGACGGCTAAGATGCTG 899
DB 361 TyrsertIleuMetThrserserLysThrGlnAspArgArgLysSerHisAsnMetPro 380
QY 900 AAGAGCTTCTGAAGGCTGCTGAGAGATG-----ATTGATTCAGTT 941
DB 381 SerSerIleleuAspAspGlyAsnLysIleGlyMetHisValSerProIleAspGlu-- 399
QY 942 ACATGAGCACTCATTTGAATGA-----CGG 971
DB 400 ---TyrSerHisPheIleAspAsnAspGluProleuTrpArgAspLysValIleProLys 418
QY 972 ACTGCTACGAGGAAGATTTTAAACCTGATGATTTGAATTTTATTCATCTGTG 1031
DB 419 IleTythrAsnGlnIleThrProThrProAspAlaIlePheAspSer----- 436
QY 1032 CAAAAGTTTCCAGGTGTTGAGACACAG-----CTGGCAAGAGGTGTG 1082
DB 437 HisLysIleTyraIleIleSerleuLeuArgIleTyraIleProGluLysArgLysPhe 456
QY 1083 TTAGAGAAACAAGCTCTGCATATGAGCGGAGCGCCCTTGCTATCCGACCTTGA 1142
DB 457 -----PheArg 458
QY 1143 GCTGCTTATGTGCTGATTAATGGCTGTGAGCCCGAATGGAATAGAGTGTG 1202
DB 459 IleTyraIleu-----ProserIlePhe 466
QY 1203 ATGAGCAAGATTTCTTGGACGA-----GAAACTACATTAGTGATGAAAACTTC 1256
DB 467 LysArgIleleuTyraIleLysPheAlaGlnIleuTyraPheMetGlnIleCysleu 486
QY 1257 GATCCTTATCCGATTAATGCTATCTCTGTTCAAAATAGTGTGGC-----ACC 1310
DB 487 GlnArgleuValIleGlnSerleuGlnIleGlnProSerleuValHisAlaIleAsn 506
QY 1311 AAGGTATTATGCAAGCGTGAAGGTTCAAGAGAAAGAACTTCAGATTAACCTTCAT 1370
DB 507 AsnleuIleLysSerSerIleGlnSerleuLys-----Asn 518
QY 1371 TGCAACAACATGACATCAAGATTAAGAGAGATTTAACTCTGTATGCCATTAAC 1430
DB 519 ValThrValThrSerAspAspLysThrSerSerGlyValIleIleleu----- 534
QY 1431 CTCCAATAGTGACCAAGACTTGGGCTTAACCTTTTCTTAACAAGAGATGAT 1490
DB 535 -----SerTyraLysGluPheLys-----SerleuSerIleuValAsn 546
QY 1491 AATACTTCTTAAGACTTTGGAGACTCATGATTACTTCCAAATCTGTCCAACTCAT 1550

```

```

DB 547 LysAspPheAsnGluProPheTrpProAsnGlnSerIleAlaAsnSer----- 562
QY 1551 GGTCAACTCAAGATGTGTGATGATCAAACTTGCACCTTTTAATGGAAGAACTCTC 1610
DB 563 -----TrpProAspPheAlaAsnLysGlnleu 571
QY 1611 CGCCAGAGAAATTCACGTGGCTTCGATTCATATTAATTTTGTGTAAGAAAT 1670
DB 572 LysArgGlyGlnIleleu-----GlnAspAlaPheAlaPheHisleuPheGlnleu-GluLe 590
QY 1671 GCCAAAGTTGCTGCTTGATCTGAATAAT 1698
DB 590 uProIleIleleuAspThrThrArgAsn 599

RESULT 11
A27421
Integrin alpha-5 chain precursor - human
N:Alternate names: CD51 antigen; serum spreading factor; vitronectin receptor alpha chain
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1989 #sequence revision 31-Mar-1989 #text change 31-Dec-2000
C:Accession: A27421; S47541; B29418; A26482; A32287; B32287; A30298; A35035
R:Suzuki, S.; Argraves, W.S.; Arai, H.; Languiño, L.R.; Pierschbacher, M.D.; Ruoslahti, I
J. Biol. Chem. 262, 14080-14085, 1987
A:Title: Amino acid sequence of the vitronectin receptor alpha subunit and comparative ex
A:Reference number: A27421; MUID:88007656; PMID:2443500
A:Accession: A27421
A:Molecule type: mRNA
A:Residues: 1-1048 <SUZ>
A:Cross-references: GB:M14648; GB:J02826; GB:M18365; NID:g340306; PIDN:AAA36808.1; PID:g3
R:Donahue, J.P.; Sugg, N.; Hawiger, J.
Biochim. Biophys. Acta 1219, 228-232, 1994
A:Title: The integrin alpha(v) gene: identification and characterization of the promoter
A:Reference number: S47541; MUID:94368864; PMID:7522056
A:Accession: S47541
A:Molecule type: DNA
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-61 <DON>
A:Cross-references: EMBL:U07375
R:Fitzgerald, L.A.; Poncz, M.; Steiner, B.; Rall Jr., S.C.; Bennett, J.S.; Phillips, D.R.
Biochemistry 26, 8158-8165, 1987
A:Title: Comparison of cDNA-derived protein sequences of the human fibronectin and vitron
A:Reference number: A30526; MUID:88163472; PMID:2450560
A:Accession: B29418
A:Molecule type: mRNA
A:Residues: 1-433 <FIT>
R:Suzuki, S.; Argraves, W.S.; Pytela, R.; Arai, H.; Krusius, T.; Pierschbacher, M.D.; Ruc
Proc. Natl. Acad. Sci. U.S.A. 83, 8614-8618, 1986
A:Title: cDNA and amino acid sequences of the cell adhesion protein receptor recognizing
A:Reference number: A26482; MUID:87041504; PMID:2430295
A:Accession: A26482
A:Molecule type: mRNA
A:Residues: 413-1048 <SUZ>
R:Cheresh, D.A.; Smith, J.W.; Cooper, H.M.; Quaranta, V.
Cell 57, 59-69, 1989
A:Title: A novel vitronectin receptor integrin (alpha-v beta-x) is responsible for discri
A:Reference number: A32287; MUID:89195223; PMID:2467745
A:Accession: A32287
A:Status: preliminary
A:Molecule type: protein
A:Residues: 31-35, 'X', 37-41 <CHR>
A:Experimental source: melanoma cell M21
A:Accession: B32287
A:Status: preliminary
A:Molecule type: protein
A:Residues: 31-33, 'X', 35, 'X', 37-41 <CHR>
A:Experimental source: lung carcinoma cell UCLA-P3
R:Blum, S.C.T.; Plov, E.F.; D'Souza, S.E.; Charest, D.A.; Prellinger III, A.L.; Ginsberg, M
J. Biol. Chem. 264, 3742-3749, 1989
A:Title: Isolation and characterization of a platelet membrane protein related to the vit
A:Reference number: A30298; MUID:89139425; PMID:2465293
A:Accession: A30298
A:Molecule type: protein

```


A:Experimental source: cv. Fribo, seed coat

C:Genetics:

A:Gene: CWIN1

C:Superfamily: beta-fructofuranosidase

C:Keywords: cell wall; glycoprotein; glycosidase; hydrolase

Alignment Scores:

Pred. No.:	2,32	Length:	575
Score:	111.50	Matches:	72
Percent Similarity:	35.71%	Conservative:	48
Best Local Similarity:	21.43%	Mismatches:	107
Query Match:	3.57%	Indels:	109
DB:	2	Gaps:	19

US-10-676-079-3 (1-1721) x T12094 (1-575)

```

QY 198 CAGGAGCCGCTGACCTGGTGAAC-----CCCTGTTCTCTGCC 236
DB 228 LYSHPSPHLSERLALYSARGTNGLYMETTRPGUCYSPROPHETRYPRO 247
QY 237 GTACACATGAGCCCACTGGCCAGGACCCGCGTCTCATCTCTGGGTTCTCA 296
DB 248 VALSERLEUGLUGLYLSANGLYLEUASP-----LEUSERMETMETGLYSNAN 265
QY 297 AAGCTTCGACCTGGCCAGAGGCTGTCT-----CCT 329
DB 266 VALLYSHIVALLLEULYSANSELEUAPRIETHRARGTYGLUTRYTHRILEGLY 285
QY 330 GCGTACCTGAGC-----TTTGGT 347
DB 286 THTTYRLEUGLHANSGLNAPRYTYRIEPROAPRYLTHSERGLUNAPRYTPLY 305
QY 348 GGCACCAACAGACACTC-----CTAATTTGCATCCCAAGAG 386
DB 306 GYLLEUARGTYRSPRYGLYSNPHERYALASERYSESPHEPHEASPROTHRYLS 325
QY 387 GATCAACCTTGAAGAGAGATTACTGG-----CAATCTCAAGTCAACG 434
DB 326 -----ASNARGARGILEETTRPGLYTPRALASNGLSERAPHRYSGLU 341
QY 435 GATATTTGCAATATGGA-----TCCATCCCTCGTATGTG----- 470
DB 342 ASPPAERVALLYLGLYTRPALAGLYLEGINALALEPROARGHRVALTRPLEUAP 361
QY 471 -----GAGGAGAAGTTACGTTGGAATGCCCTTACAGAGCAATTGCTACTCCGA 521
DB 362 SERSEARGARGGLNLEUARG--GINTPRPROVALGUGLUNLEUASNARGLEUARGGLY 380
QY 522 -----GAACACTACCAAGAAAAGTTCAAGAACAGACCTACTCAAGAGCTCTGTA 572
DB 381 LYSGLINVALGLIMELYSANRGLYLSGLYLYTYR-----LEU 396
QY 573 GATGTCCTATACACTTTTGCAAACTGCTGAGACTGAGACTTATCTTTGGCCTTAATATGCG 632
DB 397 GLUVALLYLGLYILETHRALASERGINALAPVALGLUVALTHRPHESERPHESER 416
QY 633 TTAATTAAGACAGCAATTGCAAGTGAACAGTTCTAATGCTCAAGTGTCTCTGAGACTAC 692
DB 417 LEUAPRYVALGLUVALPHEASPPROANETTRPGULASNLAGLU-----ASPLEU 433
QY 693 TGCCTTCAGAGGGGATATAACATTTCTTGGAACCTAGCAATGAACTAAGATTTCCTT 752
DB 434 CYSHLAGLNLYSGLYSERYVALARGGLYVALGLY-----PROHEGLYLEU 451
QY 753 AAGAGCTGATATTTTCAATCAATGGGCTGCAATTAGAGAAATATATATTCATATG-- 809
DB 452 THRLEUALA-----SERLYSLEYLEUGLUGLUTRYTHRSEERVALPHE 465
QY 810 -----CATTAACCTCTA-----AGAAAGTCC 830
DB 466 PHEARGVALPHELYSALALASNLVSHLSYLSILEUWECYSESERAPALAYSSER 485
QY 831 ACCTTCAAAAATGCAAAACTATATGCTCT-----GATGTT 866

```

```

DB 486 SERSELEUASNARGGLUTEUTRYLSPROSERPHEALAGLYPHEVALASNVALAPLEU 505
QY 867 GGTACGCTCCAGAAAGAACCGCTTAAGTCTGGAAGACTTCTGGAAGCTGTGGAGAA 926
DB 506 GLYASNANLVSLEYLSERLEUARGSERLEU----- 516
QY 927 GTGATGTATTCAGTTACATGCAATCATCATATTTGAATGAGCGACT 974
DB 517 ---ILEAPHLASERVALVALGLUSERPHEGLYVALGLYLYTYR 531
RESULT 13
T13425
regulatory protein K10, oocyte-specific - fruit fly (Drosophila melanogaster)
N:Alternate names: protein Eg:30B8.5
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 15-Oct-1999
C:Accession: T13425, A28826
R:Murphy, L.; Harris, D.; Barrell, B.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17668
A:Accession: T13425
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-463 <MNR>
R:Probst, E.; Deryckere, F.; Koos, C.; Haenlin, M.; Pantesco, V.; Mohler, E.
Genes Dev. 2, 891-900, 1988
A:Title: Role of the oocyte nucleus in determination of the dorsoventral polarity of Dro
A:Reference number: A28826
A:Accession: A28826
A:Molecule type: mRNA
A:Residues: 1-276, 'HH', 279-281, 'VDHR', 287-463 <PRO>
A:Cross-references: GB:X12836; NID:98148; PID:g295771
A:Gene: FLYBase:fs(1)K10
A:Cross-references: FLYBase:FBgn0000810
A:Map position: X
A:Introns: 432/3
C:Keywords: DNA binding
Alignment Scores:
Pred. No.: 2,39 Length: 463
Score: 111.00 Matches: 38
Percent Similarity: 42.86% Conservative: 10
Best Local Similarity: 33.93% Mismatches: 40
Query Match: 3.56% Indels: 24
DB: 2 Gaps: 4
US-10-676-079-3 (1-1721) x T13425 (1-463)
QY 21 CTGCGCGCAGCTGCGGGGAGAGCAAGTGAAGCCCAAGATGCTGCGTCAAG 80
DB 215 ILEAGLYGLYPHEASNGLYPHELINARGLYPRO----- 227
QY 81 CTGCGCTGCGCGCGCGCTGATGCTGCTCTGCGGCGCGTGGTCCCTCCCT 140
DB 228 PROANARGPROPRO-----ARGLEUWECGLYPROPROMETGLYPROPROMETGLYPRO 246
QY 141 GCGCGCGTCCCGCCGACCTGCGGACAGACAGACGTCGTGAGACTTGTCTGACCCAG 200
DB 247 GLY--PRO-----ARGLYPROGLYPROPROMETGLYPROG 257
QY 201 GAGCGCTGACCTGTGAGCCCTCTCTCTGCTCCGACATTGAGCCAACTGCGC- 259
DB 257 LYGLYPROTYRPROGLINMETPROPHETPROPOVALPROGLYMETARGLYPROGLYP 277
QY 260 -----CAGGACCCGCGGTTCTCATCTCTCTGCGGTTCTCAAGCTTGT 305
DB 277 ROMETGLYPROMETGLYGLYPROPROPROPROPROPROLEUAPHEARGARGAENG 297
QY 306 ACCTTGGCAAGAGCTTGTCTCTGCGTACCT 337

```

Db 297 lyProGlyProGlyPrometmetcglValPro 307

RESULT 14

T10666 hypothetical protein F6E21.40 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 16-Jul-1999 #sequence, revision 16-Jul-1999 #text, change 16-Feb-2001

C/Accession: T10666 R/Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancro

A/Reference number: 216533

A/Accession: T10666

A/Molecule type: DNA

A/Residues: 1-670 <BEV>

A/Cross-references: EMBL:AL049914; GSPDB:GN00062; ATSP:F6E21.40

A/Experimental source: cultivar Columbia; BAC clone F6E21

C/Genetic:

A/Map position: 4

A/Intons: 47/3; 87/1; 123/3; 203/3; 230/2; 255/3; 284/3; 305/1; 335/3; 347/3; 370/3; 39

C/Superfamily: Schizosaccharomyces pombe negative regulator of mitosis skbl

Alignment Scores:

Pred. No.:	2.64	Length:	670
Score:	111.00	Matches:	123
Percent Similarity:	36.36%	Conservative:	77
Best Local Similarity:	22.36%	Mismatches:	194
Query Match:	3.56%	Indels:	156
DB:	2	Gaps:	33

US-10-676-079-3 (1-1721) x T10666 (1-670)

```

QY 213 CTGTGAGCCCTGTCCTGTCCTGCATTCAGCCCACTGGCCGACCCGCG 272
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 47 LeuValAspProSerTyrArgProSerLeuValGluValAsn---GlyValAspThrGln 65
QY 273 TTCTCATCTCTCGGTTCTCCAAAGCTTGATCTGGCCGACGCTGTCTGCG 332
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66 ValLeuProValCysGlySerAspLeuVal-----LeuSerProSer 79
QY 333 TAC-----CTGAGGTTGTGGGACCAAGACAGCTTCGAAATTTTGCATCCAG--- 383
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 80 GluTrpSerSerIleValIleGlyValIleSerSerTrpIleAspLeuAspSerGluAsp 99
QY 384 -----AAGATCAACCTTTGAGAGAGAAATTACGTGCAATCTCAA 425
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100 GluValLeuArgMetAspSerGluThrThrLeuGlnGluIleAlaIleThrHis 119
QY 426 GTCAACCAAGATATTTGCAATATGATGATCCCTCTGATGGAGAGAAAGTTACGG 485
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120 LeuSerLeuGlnMetCysGlu-----ProAsp----- 128
QY 486 TTGGAATGGCCCTACCAAGAGCAATTGTACTCCGAGACATCCAGAAAAGTTCAAG 545
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 129 -----LeuThrArgProHisTyrLeuAlaGlyGlyLeu 139
QY 546 AACAGCACTTACTCAAGAGCTGTGA-----GATGTCTTATAC----- 584
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 140 ArgValSerCysCysArgSerSerPheIleSerAspGluThrPheLeuTyrIleThr 159
QY 585 -----ACTTTTGCAACTGCTCA-----GAGCTGCACTTGATCTTTGGC 623
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 160 PheAsnGlnAlaLeuThrPheCysGlySerSerLeuPheCysLeuAsnValIleSerAla 179
QY 624 CTAAATGCGTTATTAGAGACAGAGATTGCAGAGAAAGTTTATGTCTAGTGTCTC 683
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 180 LeuValSerLeuTrpLeuValIleProLeuValIle-----SerGluGlyAspSerMet 196
QY 684 CTGCACTACTGCTCTTCCAAAGGGGTATACATTTCTTGGAAGTACAGCAAT----- 724
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 197 AspAsp-----ThrSerGluGlyLeuAsnAspSerTrpGluLeuThrPheAsnSerPheArg 214

```

```

QY 735 -----GAACCTAACGTTCTCTTAAGAGCT---GATATTTTCATCAAT----- 776
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 215 LeuLeuCysGluIleHisAspSerLysLeuSerValAlaLeuAspValLeuSerThrLeuPro 234
QY 777 -----GGTCGCGAGTAGAGAGAAATTTATTCATATTCGATAACTTCTTA 821
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 235 SerGluThrSerLeuGluTyrTrpMetCysGlySer---ValArgAlaIleLeuSer 253
QY 822 AGAAAGTCACTTCAAAATGCAAACTCTATGTCTGATGTTGGTCAAGCT----- 875
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 254 ThrAspAlaPheLeuThrAsnAlaArg-----GlyTyrProCysLeu 267
QY 876 CGAAGAAAGACGGCTTAAGTCTGAAGAGCTTCCTG---AAGCTGTGTGAGAAAGGATT 932
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 268 SerLysArgHisGlnLysLeuIleAlaGlyPhePheAspHisAlaIleValValIle 287
QY 933 GATTCAGTTACATGAGCACTTACTTATTTGAATGACGAGCTGTCCAGGAGAAATTTT 992
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 288 CysGlyLysProValHisAsnLeuGlnLysProLeuAspSerSerSerGluTyrThrGlu 307
QY 993 CTAAACCTGATGTATTTGACATTTTATTTTATTCATCTGTCCAAAAGTTTCCAGGTGTT 1052
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 308 LysAsnPro-----LeuArgIleTyrLeuAspTyrValAlaTyrLeuGlnLysMet 325
QY 1053 GAGAGCAACAGCGCTGCGCAAGAGCTCTGTTAGAGAAACAAAGCTCTGCATATGAGAGC 1112
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 326 GluSerLeuSerGluGlnGluIleGluIleGluTyrArgAspPheLeuGlnAlaPro 345
QY 1113 GGAGCGCCCTGCTATCCGACACCTTGACAGCTGCTTATGTGCGTGAATAATTTGGCG 1172
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 346 LeuGlnProLeuMetCysPheAsnLeuGlnAlaGlnThrTyrGluThrPheGlu 362
QY 1173 CTGTGAGCCCGAATGGAAATGAAAGTGTGATGAGCAAGTATCTTTGAGAGAGAAAC 1232
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 363 -----ArgAspSerValLysTyrIleGlnTyrGln-----ArgAlaValGlu 376
QY 1233 TACCATTTAGTGATGATAAACTTCGATCCTTACCTGATTTATGCTATCTTCTGTTG 1292
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 377 LysAlaLeuValAspArg-----ValProAspGluLysLeuAspSerGluLeu 391
QY 1293 AAGAAATTTGTGGGACCAAGGTTTATG----- 1322
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 392 -----ThrThrValLeuMetValIleGlyAlaArgGlyProLeuVal 406
QY 1323 ---GCAAGGTGACAGTTCAAAAGAG-----AGGAACTTCGACTATACCTTCATTGC 1373
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 407 ArgAlaSerLeuGlnAlaIleGluIleThrAspArgLysLeuValTyr----- 423
QY 1374 ACAAAACAGACAAATCCAAAGCTATAAAGAGATTTAACTCTGATGCCATAAACCTC 1433
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 424 AlaValGluLysAsnProAsn-----AlaValValThrLeu 435
QY 1434 CATPAGCTACCAAG-----TACTTGGCGTTATCCCTAT 1466
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 436 HisAsnLeuValLysMetGluGlyTyrGluAspValValThrIleIleSerCysAspMet 455
QY 1467 CCTTTTCTTAC-----AAGCAAGTGATTAATCTTCTAAGACCTTTGGAGCTCAT 1520
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 456 ArgPheTrpAsnAlaProGluGlnAlaSerIleLeuValSerGluLeuGlySerPhe 475
QY 1521 GGAATTCTTCCAAATCTGTCCAACTCAATGTCTTAACCTTAAAGTGTGATGATCAA 1580
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 476 Gly-----AspAsnGlu 479
QY 1581 ACCTTGCCACTTTATG-----GAAAAACCTTCGCGCAGAGATTCACGTGAGGC 1631
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 480 LeuSerProGluCysLeuAspGlyAlaGlnArgPheLeuLysPro---AspGlyIleSer 498
QY 1632 TTGCGAGCTTTCTCATATATGTTTTTTGTC 1661
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 499 IleProSer---SerTyrThrSerPheIle 507

```

RESULT 15

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: August 28, 2004, 07:42:36 ; Search time 36.5 Seconds
(without alignments)
4910.287 Million cell updates/sec

Title: US-10-676-079-3
Perfect score: 3119
Sequence: 1 ctgagcttcgactctcg.....atactagctcgtgactgctgactg 1721

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 segs, 52070155 residues
Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+.n2p.model -DEV=slp
-Q=/cgn2.1/USFTO.spool.p/US10676079/runat_28082004_084225_23803/app_query.fasta_1.1863
-DB=SwissProt_42 -QMT=faolan -SUFFIX=isp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNIT=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10676079 @GCN 1.1.48 @runat_28082004_084225_23803 -NCPU=6 -ICPU=3
-NO_MAP -LARGEOUTER -NEG_SCORES=0 -WAIT -DSFBLCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124.5	4.0	1044	1	ITAV_MOUSE
2	112.5	3.7	314	1	PKMB_HUMAN
3	112.5	3.7	314	1	PKMB_MOUSE
4	112.5	3.6	356	1	Y670_METJA
5	112	3.6	636	1	GIDA_LACPL
6	112	3.6	688	1	YB9F_YEAST
7	112	3.6	1048	1	ITAV_HUMAN
8	111	3.6	463	1	K10_DROME
9	110.5	3.5	503	1	ANXB_RABIT
10	110	3.5	607	1	THIC_AGRIS
11	109.5	3.5	505	1	ANXB_HUMAN
12	109	3.6	627	1	SPD2_NEPCL
13	109	3.5	694	1	RP3A_HUMAN
14	108	3.5	503	1	ANXB_MOUSE
15	108	3.5	641	1	SYM_CLOTE
16	107	3.4	1247	1	CTD2_MOUSE
17	106.5	3.4	664	1	AROB_FUSNN
18	106.5	3.4	1203	1	HCN4_HUMAN

19	106	3.4	646	1	THIC_VIBRA	Q87kf0 vibrio para
20	106	3.4	1579	1	SKK2_YEAST	P53599 saccharomyc
21	105.5	3.4	1031	1	TEBT_EUPAE	O00939 euploes ae
22	105	3.4	670	1	SMW4_SCHPO	O42667 schizosach
23	105	3.4	1048	1	SRM4_RAT	O63627 rattus norv
24	105	3.4	1500	1	CPBM_HUMAN	P31327 homo sapien
25	104.5	3.4	716	1	RPR2_TAKIT	O91742 influenza a
26	104.5	3.4	788	1	PUR2_SCHPO	O20772 s bifunctio
27	104	3.3	432	1	PURA_AQUAE	O67321 aediflex aeo
28	103	3.3	796	1	PKK_CLOAB	O97363 clostridium
29	103	3.3	1586	1	SN22_HUMAN	P51531 homo sapien
30	102	3.3	481	1	CBL2_MOUSE	Q8K3m5 mus musculu
31	102	3.3	503	1	ANXB_BOVIN	P27214 bos taurus
32	102	3.4	1205	1	S122_MOUSE	P55012 mus musculu
33	102	3.3	1231	1	CFM1_HUMAN	P08603 homo sapien
34	102	3.3	1556	1	PROS_DROVI	O946a1 drosophila
35	101.5	3.3	478	1	CBL2_HUMAN	Q9CLV7 homo sapien
36	101.5	3.3	492	1	MM11_MOUSE	Q02853 mus musculu
37	101.5	3.3	495	1	PO33_MOUSE	P31361 mus musculu
38	101.5	3.3	497	1	PO33_RAT	O63262 rattus norv
39	101.5	3.3	1262	1	CA13_CHICK	P12105 gallus gall
40	101.5	3.3	1572	1	BA12_HUMAN	O60241 homo sapien
41	101	3.3	401	1	HB9_HUMAN	P50219 homo sapien
42	101	3.3	518	1	MTCO_MOUSE	Q60754 mus musculu
43	101	3.2	595	1	THIC_BACHD	Q9bj4 bacillus ha
44	101	3.2	644	1	SYM_CLOAB	O97ew5 clostridium
45	101	3.2	666	1	TKT_BACHD	Q9kad7 bacillus ha

ALIGNMENTS

RESULT 1
ITAV_MOUSE
ID ITAV_MOUSE STANDARD; PRT: 1044 AA.
AC P43406;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Integrin alpha-V precursor (Vltromectin receptor alpha subunit)
DE (CD51 antigen).
GN ITGAV.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Kidney;
RX MEDLINE=96176309; PubMed=8601592;
RA Made J., Kumar A., Liu Z., Ruoslahti E., Reichardt L., Marvaldi J.,
RA Kanwar Y.S.;
RT "Cloning of mouse integrin alphaV cDNA and role of the alphaV-related
RT matrix receptors in metastatic development.";
RL J. Cell Biol. 132:1161-1176(1996).
CC -!- FUNCTION: The alpha-V integrins are receptors for vitronectin,
CC cytoactin, fibronectin, fibrinogen, laminin, matrix
CC metalloproteinase-2, osteopontin, osteomodulin, prothrombin,
CC thrombospondin and von Willebrand factor. They recognize the
CC sequence R-G-D in a wide array of ligands. Alpha-V integrins may
CC play a role in embryo implantation, angiogenesis and wound
CC healing. Mice expressing a null mutation of the alpha-V subunit
CC gene survive until late in embryonic development and occasionally
CC even to birth. They demonstrate cleft palate, and defective
CC development of CNS and gastrointestinal blood vessels.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A
CC DISULFIDE BOND. ALPHA-V ASSOCIATES WITH EITHER BETA-1, BETA-3,
CC BETA-5, BETA-6 OR BETA-8.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: Belongs to the integrin alpha chain family.
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).

DR	EMBL; U14135; AAC52497.1; -			
DR	PIR; T10050; T10050.			
DR	HSP; P06756; IJY2.			
DR	MG; MGI:96608; Itgav.			
DR	GO; GO:0001568; P-blood vessel development; IMP.			
DR	InterPro; IPR000413; Integrin_alpha.			
DR	Pfam; PF01839; FG-GAP; 4.			
DR	Pfam; PF00357; Integrin_A; 1.			
DR	PRINTS; PR01185; INTEGRINA.			
DR	SMART; SM00191; Int_alpha; 5.			
DR	PROSITE; PS00242; INTEGRIN ALPHA; 1.			
KM	Angiogenesis; Integrin; Cell adhesion; Receptor; Glycoprotein,			
KM	Transmembrane; Signal; Repeat; Calcium.			
FT	SIGNAL	1	30	
FT	CHAIN	31	1044	
FT	CHAIN	31	885	
FT	CHAIN	887	1044	
FT	DOMAIN	31	988	
FT	TRANSLEM	989	1012	
FT	DOMAIN	1013	1044	
FT	REPEAT	46	108	
FT	REPEAT	113	178	
FT	REPEAT	179	236	
FT	REPEAT	249	302	
FT	REPEAT	303	362	
FT	REPEAT	367	427	
FT	REPEAT	432	483	
FT	CA_BIND	260	268	
FT	CA_BIND	314	322	
FT	CA_BIND	379	387	
FT	CA_BIND	443	451	
FT	SITE	1015	1019	
FT	DISULFID	89	97	
FT	DISULFID	138	158	
FT	DISULFID	172	185	
FT	DISULFID	491	502	
FT	DISULFID	508	565	
FT	DISULFID	626	632	
FT	DISULFID	698	711	
FT	DISULFID	852	900	
FT	DISULFID	905	910	
FT	CARBOHYD	74	74	
FT	CARBOHYD	290	290	
FT	CARBOHYD	296	296	
FT	CARBOHYD	615	615	
FT	CARBOHYD	704	704	
FT	CARBOHYD	835	835	
FT	CARBOHYD	851	851	
FT	CARBOHYD	869	869	
FT	CARBOHYD	941	941	
FT	CARBOHYD	956	969	
FT	CARBOHYD	976	976	
QQ	SEQUENCE	1044 AA;	115277 MW;	F06ABFF2270518BD CRC64;

Alignment Scores:

Pred. No.:	0.197	Length:	137
Score:	124.50	Matches:	104
Percent Similarity:	33.97%	Conservative:	86
Best Local Similarity:	23.10%	Mismatches:	213
Query Match:	3.99%	Indels:	184
DB:	1	Gaps:	35

US-10-676-079-3 (1-1721) X ITAV_MOUSE (1-1044)

OY		48	CCAGTAGGCCAAGAATGCTGTGGCGTCCGAAGCTGGCGGCCTGATGCTG	107
Dd		4	Prolgyl-----ArgLeuLeuLeuXpProargProlgyl-----GlyLeuLeuLeu	18
OY		108	CTGCTCCTGGGGGCGCTGGGTCCCCTCTCCCTGGCGCCTGCCCCGACCTGCAGAACGA	167
Dd		19	LeuLeuPproGlyLeuLeuLeuProLeuLa-----	28
OY		168	CAGACGTCGTGCACTGTGACTTTCTTAACCAGAGCCGCTGCACCTGTGAGCCCC---	224
Dd		29	--AspAlaPheAenLeuAsp-----ValGluSerProIagIuTyraGlyProGlu	45
OY		225	---TGATTCCGTGCCGTACCATTTGAC---GCCAACTGGCAGAGACCGCGGTTCTCTC	278
Dd		46	GlySerTyrrPheGlyPheAlaValaAspPheGluProSerThrSerArgMetPhe	65
OY		279	ATCTCTCGTGGTCTCCAAAGCTCTGACCTTGGCGCACAGAGCTTGTCTCTCGCTACCTG	338
Dd		66	LeuLeuValIGlyAlaIProLyalaAnthrThrcInProGlylleValGlu-----	82
OY		339	AGTTTGGTGGCACCAAGACAGACTTCTTAATTTTCATCCCAAGAAAGAACTT	398
Dd		83	-----GlyGlylnValleuLycyGluCyS--SerSerSerArgGcyglnProl	100
OY		399	GAAGAAGAAAGTTACTGSCAACTCTCAAGTACAACAGAGATTTTGCAATATGATTCATC	458
Dd		100	eGlubPheasBerthrGlyasn-----ArgAspTyrrAlaYsaBapPro--	115
OY		459	CCTCTGATGTGGAGAGAGTTTACGGTTGGAAATGGCCCTTACAGAGCAATTCTACTC	518
Dd		116	----LeuGluubHeysSerThrsgIntPrphcGlyAla-----	126
OY		519	CGAAGACACTACAGAAAAAGTTTCAGACAGCACCTTACAGAGACTTGTAGATGTG	578
Dd		127	-----SerValArgerLys--GlnAsp-----LyeIlleuLaCySa	139
OY		579	CT-----ATPACATTTTGCAAACTGCTCAAGACTGACCTTGANTCTTTGGCTTAAACGC	632
Dd		139	IaPoleuTyrrHstrPAyrThrGluueLysGlnGuarGluarovaIGlyThrCySp	159
OY		633	TTATTAGAACAGCAGATTGTCAGTGGAGACAGTTCTATGCTC-----T	674
Dd		159	helouGlnaerGlyThrlYserThrValGluTyrrAlaProCyBaYssetLYsaenlIeaSpA	179
OY		675	CAGTTGCTCTGGACTACTGCTCTTCCAAAGGGGTATAACATTTCTTGGGAAGTACGCAAT	734
Dd		179	IaaerGlyGlnGly-PheCysGlnGly---glyPheSerilleaPhe-----	193
OY		735	GAACCTTAACAGTTCCCTTAAGAAAGCGTCGATATTTTCAATCATGGTGTGCAATTAGAGAA	794
Dd		194	-----ThrlYsaIlaaSPryalValleuGlyGlyPro--GlySer	206
OY		795	GATTATATTCATATGCATAAATCTTCAAGAAAGCCACTTCAAAATNGCAAACCTCAT	854
Dd		207	PheTyrrTrpGln---GlyGlnleuIleSerAaspGlnValaGluIleIleSerLYeTyrr	225
OY		855	GGTCTCATGATT-----GGTCAGCCTCGAAGAAAGCGCTAGATG	896
Dd		226	AspProaenValTYrSerIleTyrrAasnandlneulaThrxarGtnrIaIGlnAla	245
OY		897	CTGAAGACTTCCGAAAGCGTGGTGGAGAAAGTATTATTCAGTTACATGCACTACAC	956
Dd		246	Ile-----PheaAspSer	250
OY		957	TATTGAATGAGCGGACTGCTACCGAGGAGATTTTCTAAACCTGATGTATGGACATT	1016
Dd		251	TyrLeu-GLyTyrrSerValaIaValaGlyAspPhe---AenGlyAspGlyIleGluAsp	268
OY		1017	TTTTATTTCACTGTGCAAAAAGTTTCCAGGTGGTGTAGAGACAC-----AGGCTTGC	1070
Dd		269	PheValaBerGlyValaProArGAlaIaIaArgThrlenuIyaeValTyrrIetyrrAaspGly	288
OY		1071	AAGAAGGTC-----TGTTNAGAGAAACAAGCTCTGCATATGAGGC--	1112

Db	QY	1113	309	QY	1122	329	QY	1179	Db	QY	1206	369	QY	1266	386	QY	1326	392	QY	1371	412	QY	1419	432	QY	1470	452	QY	1515	471	QY	1566	483	QY	1623	497	Db	PMXB_HUMAN	STANDARD	PRT	314	AA
Db	289	lysaenmetSerSerLeuHisasnPhenrhcylgluInMetalaiaIaIaTyPheGlyPhe	308	QY	1113	309	QY	1179	Db	349	leuglnaIrgalaValaIglYaapPheInThrIlyleuasnIglYpHeIgluValaPheaIa	368	QY	1206	369	QY	1326	392	QY	1371	412	QY	1419	432	QY	1470	452	QY	1515	471	QY	1566	483	QY	1623	497	Db	PMXB_HUMAN	STANDARD	PRT	314	AA
Db	289	lysaenmetSerSerLeuHisasnPhenrhcylgluInMetalaiaIaIaTyPheGlyPhe	308	QY	1113	309	QY	1179	Db	349	leuglnaIrgalaValaIglYaapPheInThrIlyleuasnIglYpHeIgluValaPheaIa	368	QY	1206	369	QY	1326	392	QY	1371	412	QY	1419	432	QY	1470	452	QY	1515	471	QY	1566	483	QY	1623	497	Db	PMXB_HUMAN	STANDARD	PRT	314	AA
Db	289	lysaenmetSerSerLeuHisasnPhenrhcylgluInMetalaiaIaIaTyPheGlyPhe	308	QY	1113	309	QY	1179	Db	349	leuglnaIrgalaValaIglYaapPheInThrIlyleuasnIglYpHeIgluValaPheaIa	368	QY	1206	369	QY	1326	392	QY	1371	412	QY	1419	432	QY	1470	452	QY	1515	471	QY	1566	483	QY	1623	497	Db	PMXB_HUMAN	STANDARD	PRT	314	AA
Db	289	lysaenmetSerSerLeuHisasnPhenrhcylgluInMetalaiaIaIaTyPheGlyPhe	308	QY	1113	309	QY	1179	Db	349	leuglnaIrgalaValaIglYaapPheInThrIlyleuasnIglYpHeIgluValaPheaIa	368	QY	1206	369	QY	1326	392	QY	1371	412	QY	1419	432	QY	1470	452	QY	1515	471	QY	1566	483	QY	1623	497	Db	PMXB_HUMAN	STANDARD	PRT	314	AA
Db	289	lysaenmetSerSerLeuHisasnPhenrhcylgluInMetalaiaIaIaTyPheGlyPhe	308	QY	1113	309	QY	1179	Db	349	leuglnaIrgalaValaIglYaapPheInThrIlyleuasnIglYpHeIgluValaPheaIa	368	QY	1206	369	QY	1326	392	QY	1371	412	QY	1419	432	QY	1470	452	QY	1515	471	QY	1566	483	QY	1623	497	Db	PMXB_HUMAN	STANDARD	PRT	314	AA
Db	289	lysaenmetSerSerLeuHisasnPhenrhcylgluInMetalaiaIaIaTyPheGlyPhe	308	QY	1113	309	QY	1179	Db	349	leuglnaIrgalaValaIglYaapPheInThrIlyleuasnIglYpHeIgluValaPheaIa	368	QY	1206	369	QY	1326	392	QY	1371	412	QY	1419	432	QY	1470	452	QY	1515	471	QY	1566	483	QY	1623	497	Db	PMXB_HUMAN	STANDARD	PRT	314	AA
Db	289	lysaenmetSerSerLeuHisasnPhenrhcylgluInMetalaiaIaIaTyPheGlyPhe	308	QY	1113	309	QY	1179	Db	349	leuglnaIrgalaValaIglYaapPheInThrIlyleuasnIglYpHeIgluValaPheaIa	368	QY	1206	369	QY	1326	392	QY	1371	412	QY	1419	432	QY	1470	452	QY	1515	471	QY	1566	483	QY	1623	497	Db	PMXB_HUMAN	STANDARD	PRT	314	AA
Db	289	lysaenmetSerSerLeuHisasnPhenrhcylgluInMetalaiaIaIaTyPheGlyPhe	308	QY	1113	309	QY	1179	Db	349	leuglnaIrgalaValaIglYaapPheInThrIlyleuasnIglYpHeIgluValaPheaIa	368	QY	1206	369	QY	1326	392	QY	1371	412	QY	1419	432	QY	1470	452	QY	1515	471	QY	1566	483	QY	1623	497	Db	PMXB_HUMAN	STANDARD	PRT	314	AA
Db	289	lysaenmetSerSerLeuHisasnPhenrhcylgluInMetalaiaIaIaTyPheGlyPhe	308	QY	1113	309	QY	1179	Db	349	leuglnaIrgalaValaIglYaapPheInThrIlyleuasnIglYpHeIgluValaPheaIa	368	QY	1206	369	QY	1326	392	QY	1371	412	QY	1419	432	QY	1470	452	QY	1515	471	QY	1566	483	QY	1623	497	Db	PMXB_HUMAN	STANDARD	PRT	314	AA
Db	289	lysaenmetSerSerLeuHisasnPhenrhcylgluInMetalaiaIaIaTyPheGlyPhe	308	QY	1113	309	QY	1179	Db	349																																

```

RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20487360; PubMed=11034547;
RA Adachi M., Browne D., Lewis B.J.;
RT "Paired-like homeobox domain proteins Phox2a/Arx and Phox2b/NBPphox have
RT similar genetic organization and independently regulate dopamine
RL beta-hydroxylase gene transcription.";
RL DNA Cell Biol. 19:539-554(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99326521; PubMed=10395799;
RA Yokoyama M., Watanabe H., Nakamura M.;
RT "Genomic structure and functional characterization of NBPphox (PMX2B) ,
RT involved in second messenger-mediated catecholaminergic cells that is
RT involved in second messenger-mediated transcriptional activation.";
RL Genomics 59:40-50(1999).
CC -1- FUNCTION: Involved in the development of several major
CC noradrenergic neuron populations, including the locus coeruleus.
CC Transcription factor which could determine a neurotransmitter
CC phenotype in vertebrates. Enhances second-messenger-mediated
CC activation of the dopamine beta-hydroxylase and c-fos promoters,
CC and of several enhancers including cAMP-response element and
CC serum-response element.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed in neuroblastoma, brain and adrenal
CC gland.
CC -1- SIMILARITY: Belongs to the paired homeobox family.
CC -1- SIMILARITY: Contains 1 homeobox domain.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/).
```

Qy	231	AACCCAGGAGATGTGAGAACCCGGGGTCCGTGCGCAGTTGGCGTAATGTGACGGACA	232
Db	221	AenGlyGlyGlyGlyGly-----GlyProSerProIa-----	221
Qy	231	GGACGAGGGGCTCAACGAGTGCAGCGGCTCCTGGGTGAAGAAGTCAGGTCACGACGT	172
		: : : : : : : : : : : : :	
Db	222	-----GlyAlaProGlyAlaAlaIaGlyProGly-----GlyPro-----	232
Qy	171	CCTGTGCTTGGCCAGAGTCCGGGCGAGGGCGCGCAAGGGAGAGCGGAGCCAGCGGGCCACGGA	112
		: : : : : : : : : : :	
Db	223	-----GlyGlyGlyProGlyLysGlyGlyAlaAlaAlaAlaAla	245
Qy	111	GCAGCAGCATCAAGCGGCGCGCGCGCAGGCTTCAGCGCAGCGCAGCAGCAGTCCTGGCTCAC	52
		: : : : : : : : : : : : :	
Db	246	Ala	265
Qy	51	CTGGCTGCTCCC	40
Db	266	AlaGlyGlyPro	269

RESULT 3	PMXB_MOUSE	STANDARD;	PRT;	314 AA.
ID	PMXB_MOUSE			
AC	O35690;			
DT	15-JUL-1999	(Rel. 38, Created)		
DT	15-JUL-1999	(Rel. 38, Last sequence update)		
DT	28-FEB-2003	(Rel. 41, Last annotation update)		
DE	Paired homeobox homeobox protein 2B (Paired-like homeobox 2B)			
DE	(PHOX2B homeodomain protein) (Neuroblastoma Phox) (NBPhox).			
GN	PHOX2B OR PMXB2B.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96040559; PubMed=9374403;			
RA	Pactyn A., Motin X., Cremer H., Goriadis C., Brunet J.-F.;			
RT	"Expression and interactions of the two closely related homeobox			
RT	genes Phox2a and Phox2b during neurogenesis.";			
RL	Development 124:4065-4075 (1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99326521; PubMed=10395798;			
RA	Yokoyama M., Watanabe H., Nakamura M.;			
RT	"Genomic structure and functional characterization of NBPhox (PMXB2B),			
RT	a homeodomain protein specific to catecholaminergic cells that is			
RT	involved in second messenger-mediated transcriptional activation.";			
RL	Genomics 59:40-50 (1999).			
CC	-1- SUBCELLULAR LOCATION: Nuclear (By similarity).			
CC	-1- SIMILARITY: Belongs to the paired homeobox family.			
CC	-1- SIMILARITY: Contains 1 homeobox domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Y14493; CAA74833.1; -.			
DR	EMBL; AB015672; BAA82671.1; -.			
DR	HSSP; P06601; 1FTJ.			
DR	TRANSPAC; T03976; -.			
DR	MED; MGI:1100882; Phox2b.			
DR	InterPro; IPR001356; Homeobox.			
DR	InterPro; IPR007104; Paired_homeo.			
DR	Pfam; PF00046; homeobox; 1.			
DR	ProDom; PD000010; Homeobox; 1.			
DR	SMART; SM00389; Hox; 1.			
DR	PROSITE; PS00027; HOMEBOX_1; 1.			
DR	PROSITE; PS50071; HOMEBOX_2; 1.			

KW	Homeobox, DNA-binding; Developmental protein; Nuclear protein.
FT	DNA_BIND 98 157 HOMEOBOX.
FT	DOMAIN 159 167 POLY-ALA.
FT	DOMAIN 212 217 POLY-GLY.
FT	DOMAIN 241 260 POLY-ALA.
SO	SEQUENCE 314 AA; 31621 MW; 40737P71948B595A CRC64;

Alignment Scores:	
Pred. No. :	1.1
Score :	112.50
Percent Similarity:	46.43%
Best Local Similarity:	40.48%
Query Match:	3.73%
DB:	1
Length:	314
Matches:	34
Conservative:	5
Mismatches:	20
Indels:	25
Gaps:	4

US-10-676-079-3 (1-1721) x PMXB_MOUSE (1-314)

[illegible]

RESULT 4	x670.METUA		
ID	_x670.METUA	STANDARD;	PRT; 356 AA.
AC	Q58084;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Hypothetical protein M0670.		
GN	M0670.		
OS	Methanococcus jannaschii.		
OC	Archaea; Euryarchaeota; Methanococci; Methanococcales;		
OC	Methanocaldococcaceae; Methanocaldococcus.		
OX	NCBI_TaxID=2190;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=JAL-1 / DSM 2661 / ATCC 43067;		
RX	MEDLINE=96337999; PubMed=8688087;		
RA	Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,		
RA	Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,		
RA	Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,		
RA	Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,		
RA	Scott J.L., Geoghagen N.S.M., Wetland J.F., Fuhrmann J.L., Nguyen D.,		
RA	Uutterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,		
RA	Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Bordovsky M.,		
RA	Klank H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;		
RT	"Complete genome sequence of the methanogenic archaeon, Methanococcus		
RL	jannaschii".		
RL	Science 273:1058-1073(1996).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		

```

CC -----
DR EMBL, U67514; AAB98664.1; -.
DR PIR, F64383; F64383.
DR TIGR, M06670; -.
DR InterPro, IPR008471; DUF752.
DR InterPro, IPR000051; SAM_bind.
DR Pfam, PF05430; DUF752; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 356 AA; 41683 MW; D7B8BA2E16A92E11 CRC64;

Alignment Scores:
Pred. No.: 1.14 Length: 356
Score: 112.50 Matches: 85
Percent Similarity: 33.17% Conservative: 48
Best Local Similarity: 21.20% Mismatches: 143
Query Match: 3.61% Indels: 125
DB: 1 Gaps: 18

US-10-676-079-3 (1-1721) x Y670_METUA (1-356)

QY 438 ATTTCGAATAT-----GGATCCATCCCTCTGATGTGGAGGAATTGA 482
DB 14 IleaIglYsIYrMeIlyleIYrAsngIlybAsngIulYsAspIleYsgIulrIglu 33
QY 483 CGGTTGGAATGGCCCTTACCGAGGACAAATTCTACTCCGAAACACTACCGAAGAAAGTTC 542
DB 34 IleySglu-----LeuYsGluGluIhIbValIleuValIgluThr 46
QY 543 AAGAACAGACCTTACTCAAGAAAGCTCTGTAGAT----- 575
DB 47 GluaSpGlyThrYrThrleuYsAlaGluAspGluGluGluMeIserIysVal 66
QY 576 -----GTGCTACACTTTTGCAACCTGCTGACGACTGACCTGATCTT 620
DB 67 GlYAlaLeuYsGluAlaIleIYrIysPheAlaIYsProSer----- 80
QY 621 GGCTTAATGCGTTATTAGAACAGACAGATTGTCAGTGAACAGTTCTATGCTGAGTTG 680
DB 81 -----LysIleThrAspLeu-----SerAsnProArg--- 89
QY 681 CTCCTGCACTACTGCTCTTCCAGGGGTATTAACATTCTTGGGAACCTAGCAATGAACCT 740
DB 90 ValIleuAspLeuCySerGIYmeGIYrAsnAlaIleAlaIleuIhIstYrAsnYs 109
QY 741 AACAGT-----TTCCCT 752
DB 110 AsnAlaGluIleAspMeValIgluIleCySgluGluValIleuPheIleuThrIleuPheIleu 129
QY 753 -----AAGAAAGCTGATATTTTCATCAAT 776
DB 130 AspIleProIYrIYsGluIhIsgIuIleIleIYsAspIYsValAlrGluIYrIhIleuAsn 149
QY 777 GGGTGGCACTTAGGAGAAATTATTAATTCATTGAATTAACCTTAAGAAAGTCCACTTC 836
DB 150 -----LysIleGlyIYlEgluIYr-----LysSerAspIYr 159
QY 837 AAAAATGCAAACTCTATAGTCTCTGATGTTGGTCAGCTCGAAGAAAGCGCTAAGATG 896
DB 150 AspaAnIleAsnLeuIYr-----ValGIYAspAlaAlrYlYsPheIleIleYsSer 176
QY 897 CTGAAGAGCTCTCTGAGGCTGTGGAGAGATGATTGATTCATGATGATGATCACTAC 956
DB 177 AspYlYsIYr-----AsnValValPheIhIAspAla 187
QY 957 TATTTGAATGACGAGCTGCTAAC-----AGGGAAGATTTTCTAAACCTGATGATTTG 1010
DB 188 PheSerProIYsArgAspProThrIleuIYrThrIYrAspPheIleu----- 202
QY 1011 GACATTTTATTTCACTCTGCAAAAAGTTTCCAGAGTGTGAGAGACACAGCCTGGC 1070
DB 203 -----LysGluIleIYrIYsArgMeCIluAspAn-----Gly 213
QY 1071 AAGAAAGTCTGTGTTAGAGAAACAAGCTCTGCAATGAGAGCGAGCGCCTTGCTATCC 1130

```

```

DB 214 ValIleuIle-----SerYrSerSerAlaIleProPheArgSer 226
QY 1131 GACACCTTTCGACGCTGCTTATATGCTGATTAATTTGGGCTGTCAGCCCGAATGGGA 1190
DB 227 AlaIleuValAspCySgIYpPheValIleSerGluYsGluSerValGIYArGlyeArGly 246
QY 1191 ATGAAAGGTGATGAGGACCAAGTATTTCTTGACAGGAAATCCATTATGTGATGATA 1250
DB 247 IleThrIleuAlaIYrIYsAsnProAsnIhIeYsProAsnAlrIleAsnGluValAspIlu 266
QY 1251 AACTTCGAT-----CCTTACCTGATTTATGGCTATCTTCTGT 1289
DB 267 ArgValIleAlaIleuSerValIleAlaIleuProIYrArgAspGluThrIleuSerIeuthr 286
QY 1290 TTCAGAAATTTGGTGGGACCAAGGTTAATGGCAAGCTGTCAGAGTTCAAGGAAGAG 1349
DB 287 LysAspIYsIleIleGluAspArgIgluIuArgIgluYsLeuYsGluIYsIleuIle 306
QY 1350 AAGCTTCAGATATACCTTCATTGACCAAAACACTGACATCCAGGTATTAAGAAAGAGAT 1409
DB 307 LysIleGlyIYsIYrIleuSerThrIYsGluIleYsIYsGlyAsnIleProGluIuIle 326
QY 1410 TTAACCTGTATGCC-----ATAAACCTCCATTAACGTACCAAGTACTGCGTTACC 1463
DB 327 LeuYsIleGluIYsGluAspLeuAsnSerSerGluIleIleYsIYsMeIArgIleuYs 346
QY 1464 TRT 1466
DB 347 Phe 347

RESULT 5
GIDA_LACPL STANDARD; PRT; 636 AA.
ID GIDA_LACPL
AC Q88XK6;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Glucose inhibited division protein A.
GN GIDA OR LP 3681.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillius.
OX NCBI_TaxId=1590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 8826 / WCFS1;
RX MEDLINE=22480296; PubMed=12566566;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hotter S.M., Nierop Groot M.N., Kerkhoven R., De Vries W., Ursing B.,
RA De Vos W.M., Stezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
CC - FUNCTION: Not known.
CC - SIMILARITY: Belongs to the gida family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AL935262; CAD65746.1; -.
DR HAMAP, MF 00129; -.
DR InterPro, IPR001327; FAD_pyr_redox.
DR InterPro, IPR002218; GIDA.
DR InterPro, IPR000205; NAD_BS.
DR InterPro, IPR001100; Pyr_redox.
DR Pfam, PF01134; GIDA; 1.

```

DR PRINTS; PR00368; EADPNR.
 DR PRINTS; PR00411; EMDRDTASEI.
 DR PROSITE; PS01280; GIDA_1; 1.
 DR PROSITE; PS01281; GIDA_2; 1.
 KW Complete proteome.
 SQ SEQUENCE 636 AA; 70735 MW; 5A521AF0EF3AB93A CRC64;

Alignment Scores:

Score:	1.46	Length:	636
Percent Similarity:	112.00	Matches:	75
Best Local Similarity:	34.17%	Conservative:	48
Query Match:	20.83%	Mismatches:	111
	3.59%	Indels:	126
		Gaps:	18

US-10-676-079-3 (1-1721) x GIDA_LACPL (1-636)

```

QY 42 GAGCAGCAGAGTGAAGCCAGAGATGCTGCGCTCGAAGCTGCGCTGCGCGCGCGCTG 101
   |||||
DB 222 G|U|G|I|P|T|G|I|Y|A|P|P|G|I|-----
   |||||
QY 102 ATGCTGCTGCTGCGCGCGCGCTGCGCTGCTGCGCGCGCGCTGCGCGCGCTGCG 161
   |||||
DB 229 -----ProthiA|P|H|E|S|E|R|P|H|E|G|I|U|T|H| 236
   |||||
QY 162 CAAGCAGCAGCAGCTGCTGAGCTGGAAGCTTC-----TTCAGCCAG 200
   |||||
DB 237 L|Y|A|A|P|A|A|P|T|Y|T|I|E|A|P|L|E|U|Y|E|N|I|G|I|N|E|U|S|E|C|Y|T|T|P|L|E|U|T|H|T|Y|T|H|A|N 256
   |||||
QY 201 GAGCCGCTGAGCAGCTGAGCAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 260
   |||||
DB 257 G|U|T|H|T|H|T|H|Y|L|Y|V|E|I|L|E|-----I|E|A|T|G|U|A|N|L|E|U|A|P 268
   |||||
QY 261 AGCGAGCCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 320
   |||||
DB 269 A|Y|G|A|A|P|A|P|A|P|A|P|A|P|A|P|A|P|A|P|A|P|A|P|A|P|A|P|A|P 279
   |||||
QY 321 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 356
   |||||
DB 280 V|A|I|G|I|P|T|G|I|Y|C|Y|P|P|O|S|E|R|I|E|G|I|U|A|P|L|Y|I|E|V|A|I|A|P|H|E|A|A|P|L|Y|E|Y|S 299
   |||||
QY 357 ACAGACTTCTTAATTTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 416
   |||||
DB 300 A|Y|H|I|G|I|N|E|U|P|H|E|U|G|I|P|T|G|I|Y|A|Y|G|A|N|T|H|A|P|G|I|U| 315
   |||||
QY 417 CAATCTCAAGTCAACAGAGATTTGCAATATGGA-----TCCATCCCTCTGAT 467
   |||||
DB 316 T|Y|V|A|I|G|I|N|-----G|I|Y|E|U|S|E|R|T|H|S|E|R|P|T|G|I|U|G|I|U 327
   |||||
QY 468 GTGAGAGAGAGATTA-----CCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 482
   |||||
DB 328 V|A|I|G|I|N|A|P|T|G|I|Y|E|U|H|S|E|R|I|E|Y|G|I|U|E|U|G|I|U|A|P|A|G|I|U|E|T|C|A|P|P|O 347
   |||||
QY 483 -----CGTGGAGATG-----CCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 524
   |||||
DB 348 G|I|Y|T|A|I|L|E|G|I|U|Y|A|P|V|A|I|A|A|P|T|O|T|Y|I|N|-----L|E|U|Y|E|A| 353
   |||||
QY 525 C|A|T|T|A|C|A|G|A|A|A|G|T|T|C|A|A|G|A|C|A|G|A|C|A|G|A|C|T|G|T|G|A|T|G|T|G|T|A|C 584
   |||||
DB 364 T|H|L|E|U|G|I|U|T|H|Y|S|E|U|V|A|I|Y|A|N|L|E|U|T|Y|T|H|A|I|G|I|N|T|H|A|N|G|I|Y|T|H|S|E|R 383
   |||||
QY 585 A|C|T|T|T|G|C|A|A|C|T|G|C|A|G|A|C|T|G|A|T|T|T|G|C|T|T|G|C|T|T|A|A|T|G|C|T|T|A|A|G|A|C|A 644
   |||||
DB 384 G|I|Y|T|C|I|G|I|U|A|I|A|A|G|I|G|I|U|L|E|U|I|E|A|I|G|I|Y|I|E|A|N|A|I|G|I|Y|L|E|U|A|G|A|A 403
   |||||
QY 645 G|C|A|G|A|T|-----T|T|G|C|A|T|G|A|A|C|A|G|T|T|A|T|G|T|C|A|G|T|G|T|C|T|G|A|C 689
   |||||
DB 404 L|E|U|A|P|A|P|G|I|U|G|I|N|P|H|E|T|H|L|E|U|Y|A|Y|G|S|E|R|A|P|A|T|Y|T|I|E|G|I|Y|A|I|E|U|T|I|E|A|P 423
   |||||
QY 690 T|A|C|T|G|C|T|T|C|A|G|G|G|T|A|C|A|T|T|T|G|G|A|C|T|A|----- 728
   |||||
DB 424 A|P|L|E|U|V|A|I|H|T|H|Y|S|E|U|G|I|U|P|T|O|T|Y|T|A|Y|G|L|E|U|E|U|T|H|S|E|R|A|P|A|I|A|G|I|U|T|Y 443
   |||||

```

```

QY 729 -----GGCAGAGACCTTAACAGCTTCTTAAGAGCGCTGATTTTCATCAATGGG 779
   |||||
DB 444 A|Y|G|I|U|L|E|U|A|Y|H|I|S|A|P|A|N|A|A|P|L|E|U|A|Y|G|L|E|U|T|H|A|P|-----L|Y|S|I|Y 460
   |||||
QY 780 TCCAGCTAGAG-----GAAGTATTAATTCATTAATTCATTAATTCATTAATTCATTAAG 827
   |||||
DB 461 A|Y|G|I|U|L|E|U|G|I|Y|L|E|A|P|A|P|G|I|U|A|Y|T|A|I|A|A|P|H|E|G|I|U|-----A|A|Y|U|A|Y|G 478
   |||||
QY 828 TCCAGCTTCAAAATTCAGAACTATGCTGCTGATGTTGTCAGCGCTGAGAGAAAGACG 887
   |||||
DB 479 G|I|N|A|I|L|E|Y|A|N|G|I|U|E|U|-----A|P|A|Y|G|L|E|U|G|I|Y|L|E|A|Y|I|L|E|U|S|P|P|O 495
   |||||
QY 888 GCTAAGATCTGAGAGCTTCTTGAAGCT-----GGTGAAGAGTGAATTCATTCAGTT 941
   |||||
DB 496 A|P|L|E|U|P|G|I|U|A|A|N|A|A|P|H|E|U|Y|G|A|I|N|A|N|S|E|R|G|I|Y|L|E|U|Y|S|A|P|G|I|Y|A|I 515
   |||||

RESULT 6
YB9F_YEAST
ID YB9F_YEAST STANDARD; PRT; 688 AA.
AC P38338;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUL-1998 (Rel. 35, Last annotation update)
DE Hypothetical 80.4 kDa protein in POP4-SH1 intergenic region.
GN YBR259W OR YBR1727.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288c;
RX MEDLINE=93220397; PubMed=8465606;
RA Doignon F., Biteau N., Crozet M., Aigle M.;
RT "The complete sequence of a 19,482 bp segment located on the right
   arm of chromosome II from Saccharomyces cerevisiae.";
RL Yeast 9:189-199(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sb.ch).
CC -----
DR EMBL; X70529; CAA85923.1; -.
DR EMBL; Z36128; CAA85222.1; -.
DR PIR; S32961; S32961.
DR GenBank; J38802; -.
DR SCD; S0000463; YBR259W.
KW Hypothetical protein.
SQ SEQUENCE 688 AA; 80426 MW; 0BA84837BD7A4B30 CRC64;

Alignment Scores:
Pred. No.: 1.5 Length: 688
Score: 112.00 Matches: 101
Percent Similarity: 35.10% Conservative: 78
Best Local Similarity: 19.80% Mismatches: 168
Query Match: 3.59% Indels: 163
Gaps: 25

US-10-676-079-3 (1-1721) x YB9F_YEAST (1-688)
QY 438 ATTGCAATATGAGTCAATCCCTCTGATGTGAGAGAGAGATTAAGCTTGAATGGCC 497
   |||||
DB 164 M|E|T|A|I|G|I|U|Y|S|E|R|T|Y|T|P|A|P|S|E|R|A|P|L|Y|A|Y|G|I|N|E|U|G|I|N|P|H|E|T 183
   |||||
QY 498 T|A|C|A|G|A|C|A|T|T|G|C|A|G|A|C|A|T|C|A|G|A|C|A|T|C|A|G|A|C|A|C|T|A|C 557
   |||||
DB 184 T|Y|G|I|U|P|H|A|Y|G|I|U|L|E|U|Y|S|E|U|V|A|I|Y|P|H|E|T|Y|C|I|U|S|E|R|P|H|E|A|P|L|E|U 203
   |||||
QY 558 TCAAGAGCTCT-----GTAGATGTGCTATACACT 587
   |||||

```

```

Db      204 GlnlySerSerAspProleuylsGluLeuIleIleProtrpGluLysIleValTyrVal 223
Qy      588 TTGGCAACTGCTGAGACTGAGCTTGATCTTGGCTTAAGCCTTATTAAAGACAGA 647
Db      224 ---AlaAsnCyS-----IleAspAlaPheThrGlyGluGlnValArgIleAspGlyAla 240
Qy      648 GATTGGACGTGAGACAGTCTTAAT-----GCTCAGTTGGCTC 683
Db      241 GluLeuIleIleTrpThrSerLysAsnLeuValIlePheSerSerIleSerSerAlaValLeuArg 260
Qy      684 CTGAC-----TACTGCTCTCCAGGGGAT----- 710
Db      261 LeuAsnAspLeuGlnAsnMetPheSerAlaPheArgProTyrGlyGluGlnAlaLeuVal 280
Qy      711 -----AACATTCTTGGGAAGTGGCAAGTGAAGACCTTAACAGTTTC 749
Db      281 GlnAspPheAlaHisIleArgSerLeuLysTrpAspSerAsnAspLysValGluSerLeu 300
Qy      750 CTTAAGAAGGCT-----GATATTTTC-----ATCAATGGGCTGAGTTAGGA--- 791
Db      301 IleArgAlaLeuIlePheAsnAspMetPheProTyrPheAsnLysGluGlnValAspThr 320
Qy      792 ---GAAGATTATATTCATTTGATTAACCTTGAAGAG----- 827
Db      321 LysAlaAspGlyIlePhePheLeuArgLeuLeuArgLysAsnPheLysGluHisIleAsn 340
Qy      828 -----TCCACTTCAAAAT--- 842
Db      341 AspValLysAspPheHisIleGlnValIleLysTyrLeuAsnSerGlnPheLysAsnAsn 360
Qy      843 ---GCAAACTCATGCTGCTGATGTGTGCTGAGCTCGAAGAAAGCGCTGAAGATGCTG 899
Db      361 TyrSerThrLeuMetThrSerSerLysThrGlnAspArgLysSerHisAsnMetPro 380
Qy      900 AAGAGCTTCTTGAAGGCTGAGAGAGT-----ATTGATTCAGTT 941
Db      381 SerSerIleLeuAspArgLysAsnLysIleGlyMetHisValSerProIleAspGlu--- 399
Qy      942 ACATGGCATCATCTATTGTAATGA-----CGG 971
Db      400 ---TyrSerHisPheIleAspAsnAspGluProLeuTrpArgAspLysValTyrProLys 418
Qy      972 ACTGCTACACGAGGAGATTTCTTAACCTGATGTGATGAGACATTTTATTCATGTG 1031
Db      419 IleTyrThrAsnGluGlnThrProThrProAspAlaSerAlaIlePheAspSer----- 436
Qy      1032 CAAAAGTTTCCAGGTGTTGAGACACAGG-----CTTGGCAAGAGTCTG 1082
Db      437 HisLysIleTyrAlaIleIleSerLeuLeuArgTyrTyrLeuProLysArgLysPhe 456
Qy      1083 TTAGGAGAAACAAGCTCTGCATATGAGGCGGAGCGCTTGTATCCAGACCTTTGCA 1142
Db      457 -----PheArg 458
Qy      1143 GCTGCTTATGTCGTGATAAATTGGGCTGTACCCGAAATGGGAATAGAAGGCG 1202
Db      459 IleTyrTyrLeu-----ProSerIlePhe 466
Qy      1203 ATGAGGCAAGTATCTTTGAGCA-----GGAACTACCATTTAGTGAAGAACTTC 1256
Db      467 LysArgIleLeuTyrTyrGlyAlaLysPheAlaGlnLeuTyrPheMetGluGlyCysLeu 486
Qy      1257 GATCTTTACCTGATTATTTGGCTATCTCTCTTCAAGAAATGGTGGC-----ACC 1310
Db      487 GluArgLeuValIleGluSerLeuGlnIleLeuGluProSerLeuValHisAlaIleAsn 506
Qy      1311 AAGGTATTAATGCGACGCGCAAGGTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1370
Db      507 AsnLeuIleLysSerSerIleGluSerLeuLys-----Asn 518
Qy      1371 TGCACAAACATGCATCAATCCAAAGTTAAAGAGAGATTTAACTCTGTATGCCATTAAC 1430

```

```

Db      519 ValThrValThrSerAspAspLysThrSerSerGlyValIleIleLeu----- 534
Qy      1431 CTCATTAACGTACCAACTACTGGCGTTACCCATACCTTTTCTTAACAGCAAGTGAT 1490
Db      535 -----SerTyrLysGluPheLys-----SerLeuSerGluValAsn 546
Qy      1491 AAATACCTTCTAAGACTCTGGGACCTTCATGATTAATCTTCCAAATCTGTCCAACTCAAT 1550
Db      547 LysAspPheAsnGluProPheTrpProAsnGlnSerIleAlaAsnSer----- 562
Qy      1551 GGTTAACCTTAAGAGAGTGGATGATCAAACTTGGCCACCTTTAATGAAAGAACTTTC 1610
Db      563 -----TrpAspPheAlaAsnLysGlnLeu 571
Qy      1611 CGCCGAGAGATTCACGTGGGCTTGCACCTTTCATATATTTTGTGATAGAAGAT 1670
Db      572 LysArgGlyGlnIleLeu---GlnAspAlaPheAlaPheHisLysPheGluIle-GluLe 590
Qy      1671 GCCAAGTTGCTGCTTGATCTGCAATCTGAAGAT 1698
Db      590 uProIleIleIleAspThrThrArgAsn 599

RESULT 7
ID ITAV HUMAN STANDARD; PRT; 1048 AA.
AC P06756;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Integrin alpha-V precursor (Vltronectin receptor alpha subunit)
DE (CD51 antigen).
GN ITGAV OR VNRA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP MEDLINE=88007656; PubMed=2443500;
RX Suzuki S., Argaves W.S., Arai H., Languino L.R., Pierschbacher M.D.,
RA Ruoslahti E.;
RT "Amino acid sequence of the vltronectin receptor alpha subunit and
RT comparative expression of adhesion receptor mRNAs.";
RL J. Biol. Chem. 262:14080-14085(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20422500; PubMed=10965141;
RA Sims M.A., Field S., Barnes M.R., Shaikh N., Ellington K.,
RA Murphy K.B., Spurr N.K., Campbell D.A.;
RT "Cloning and characterisation of ITGAV, the genomic sequence for human
RT cell adhesion protein (vltronectin) receptor alpha subunit, CD51.";
RL Cytogene. Cell Genet. 89:268-271(2000).
RN [3]
RP SEQUENCE OF 413-1048 FROM N.A.
RX MEDLINE=87041504; PubMed=2430295;
RA Suzuki S., Argaves W.S., Pytela R., Arai H., Krusius T.,
RA Pierschbacher M.D., Ruoslahti E.;
RT "CDNA and amino acid sequences of the cell adhesion protein receptor
RT recognizing vltronectin reveal a transmembrane domain and homologies
RT with other adhesion protein receptors.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:8614-8618(1986).
RN [4]
RP SEQUENCE OF 31-41.
RX MEDLINE=89195223; PubMed=2467745;
RA Chersesh D.A., Smith J.W., Cooper H.M., Quaranta V.;
RT "A novel vltronectin receptor integrin (alpha v beta x) is
RT responsible for distinct adhesive properties of carcinoma cells.";
RL Cell 57:59-69(1989).
RN [5]
RP CARBOHYDRATE-LINKAGE SITE ASN-615.
RX MEDLINE=22660472; PubMed=12754519;
RA Zhang H., Li X.-J., Martin D.B., Aebersold R.;
RT "Identification and quantification of N-linked glycoproteins using

```


FT TURN 195 196
 FT STRAND 200 203
 FT TURN 204 205
 FT HELIX 206 209
 FT TURN 210 210

Alignment Scores:

Pred. No.: 1.69 Length: 1048
 Score: 112.00 Matches: 112
 Percent Similarity: 35.95% Conservative: 62
 Best Local Similarity: 23.14% Mismatches: 171
 Query Match: 3.59% Indels: 139
 DB: 1 Gaps: 27

US-10-676-079-3 (1-1721) x ITAV_HUMAN (1-1048)

QY 93 CCGCGCGTGAATCTGTCTCTCTGGGCGCGTGGT---CCCTCTCTCCCTGGCGCGT 149
 DB 4 ProProArgArgArgLeuArgLeuGlyProArgGlyLeuProLeuLeuSerGlyLeu 23
 QY 150 CCGCGCGTGGCGGACAGACAGAGCGTGGACCTTGACCTTTCACCGAGAGCGGTG 209
 DB 24 LeuLeuProLeuCyArgAlaPheAsnLeuAspValAsp-----SerProAla 39
 QY 210 CACCTGTGTAGCGCC---TCGTTCCTGCTCCGTCACCATTCAGCGCAACCTG---GCC 260
 DB 40 GluTyrseryGlyProGluGlySerTyRheGlyPheAlaValAspPheValProSer 59
 QY 261 ACGAAGCCCGGTTCTCTACCTCTGCGGTTCTCAAGAGCTTGACCTTGCGCAGAGGC 320
 DB 60 AlaSerSerArgMetPheLeuLeuValGlyAlaProGlyAlaAsnThrThrGlnProGly 79
 QY 321 TTGTCTCCCGGTACCTGAGTTTGTTGGCCACAGACAGACTCTCTATTTTCATGCC 380
 DB 80 ILeValGlu-----GlyGlyGlnValLeu-LysCyAspTrpSerSerTh 94
 QY 381 AAGAAGATCAACCTTTGAGAGAGAGAGTGTGCAATCTCAAGTCAACAGAGATAT 440
 DB 94 rArgTrgGysGlnProLeuGluPheAspAlaThrGlyAsn-----ArgAspTy 110
 QY 441 TGCAGATATGATGCCATCCCTCTGTATGTGAGAGAGAGATGAGGTTGGAATGGCCCTAC 500
 DB 110 rAlaLysAspAspPro-----LeuGlnPheLysSerHisGlnTrpPheGlyAlaSe 127
 QY 501 C---AGAGACAT-----TGTACTCCGAGAGACATAC 530
 DB 127 rValArgSerLysGlnAspLysIleLeuAlaCysAlaProLeuTyRHisTrpArgThrG 147
 QY 531 CAGAAAAGTTCAGAACAGACACTTACAGAAAGCTGTAGATGTGCTATACACTTT 590
 DB 147 uMetLysGlnGluArg-----GluProValGlyThrCyPhe----- 159
 QY 591 GCAAACTGCTCAGAGACTGTGATCTTTGGCTTAATGCTTATTAAGAACAGACAT 650
 DB 160 -----LeuGlnAspGlyThrLysThrValGlu--TyRAlaProCyArgSerGlnAsp 176
 QY 651 TTGCACTGGAACAGTTCTATGCTCAGTTGCTCTGAGTACTGCTCTTCCAAAGGCTAT 710
 DB 177 ILeAspAlaAspGlyGlnGly-----PheCySerGlnGly--GlyPhe 189
 QY 711 AACATTTCTTGGAACTAGCAATCACTAACAGTTCTTAAAGAGGCTGATATTTTC 770
 DB 190 SerIleAspPhe-----ThrLysAlaAspArgVal 199
 QY 771 ATCAATGGGTCCGAGTGAAGAGATTAATTCATTCATTAACCTTAAAGAAAGTCC 830
 DB 200 LeuLeuGlyGlyPro---GlySerPheTyRTrpGln---GlyGlnLeuLeuSerAspGln 217
 QY 831 ACCTTCAAAAATGCAAAAATCTATGCTCTGATGTT-----GCTCG 872
 DB 218 ValAlaGlnIleValSerLysTyRAspProAsnValTySerIleLysTyRAsnGln 237
 QY 873 CCTCGAAGAAAGCGGCTAAGATGTGAAGAGCTTCTGAAGGCTGTGAGAAATGAT 932

DB 238 LeuAlaThrArgThrAlaGlnAlaIle----- 246
 QY 933 GATTGATTACATGGACATCATTCTATTTGAATGACCGAGCTGCTACACGGAAGATTTT 992
 DB 247 -----PheAspAspSerTyRLeu--GlyTyRSerValAlaValGlyAspPhe 261
 QY 993 CTAAACCTGATGTGTGACATTTTATTCATGCTGCAAAAAGTTTCCAGGTGGTT 1052
 DB 262 ---AanGlyAspGlyIleAspAspPheValSerGlyValProArgAlaAlaArgThrLeu 280
 QY 1053 GAGACGACC-----AGGCTGGCAAGAGGTC-----TGTTAGAGAA 1091
 DB 281 GlyMetValTyRLeTyRAspGlyLysAsnMetSerSerLeuTyRAsnPheThrGlyGlu 300
 QY 1092 ACAAGCTGTGATATGAGAGC----- 1112
 DB 301 GlnMetAlaIleTyRPhenGlyPheSerValAlaAlaThrAspIleAsnGlyAspAspTyR 320
 QY 1113 -----GAGCGCGCTTGCTATCCGACACTTGTGACGTGCG---TTTATG 1154
 DB 321 AlaAspValPheIleGlyAlaProLeuPheMetAspArgGlySerAspGlyLysLeuGln 340
 QY 1155 TGGCTGATTAATTGGCGCTGTACGCCGAATG----- 1187
 DB 341 GluValGlyGlnValSerValSerLeuGlnArgAlaSerGlyAspPheGlnThrThrLys 360
 QY 1188 -----GGAATRGAAGTGTGTATGAGCAAGATATTTTGACAGCAGAACTACCATTTA 1241
 DB 361 LeuAsnGlyPheGlnValPheAlaArg-----PheGlySerAlaIleAlaProLeu 377
 QY 1242 GTGATGAAGAACTTGATCCTTATTCATTTATGCTATGCTTCTGTCAAGAAATGG 1301
 DB 378 GlyAspLeuAspGlnAspGlyPheAsnAsp----- 387
 QY 1302 GTGGGACCAAGGTGTTAATGCGAAGCGTGCAAGGTTCAAGAGAGAGAGCTTGAGTA 1361
 DB 388 -----IleAlaIleAlaIleAlaIleProTyRGlGlyGluAspLysLeuGlyIleVal 403
 QY 1362 TRCCTT 1367
 DB 404 TyRile 405

RESULT 8
 ID_K10_DROME STANDARD; PRT: 463 AA.
 AC P13468; O46075; Q9W505;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE DNA-binding protein K10 (Female sterile protein K10).
 GN FS(1)K10 OR EG:30B8.5 OR CG3218.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidae; Drosophilidae; Drosophila.
 OX NCBI_Taxid:7227;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Prost E., Deryckere F., Roos C., Haenlin M., Pantescio V.,
 RA Mohler V.;
 RT "Role of the oocyte nucleus in determination of the dorsoventral
 RT polarity of Drosophila as revealed by molecular analysis of the K10
 RT gene";
 RL Genes Dev. 2:891-900(1988).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE=20196011; PubMed=10731137;
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
 RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
 RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,

RT "Molecular cloning of rabbit CAP-50, a calyculin-associated annexin protein.";
 RL Biochem. Biophys. Res. Commun. 186:1227-1235(1992).
 CC -1- FUNCTION: Binds specifically to calyculin in a calcium-dependent manner.
 CC -1- DOMAIN: A pair of annexin repeats may form one binding site for calcium and phospholipid.
 CC -1- SIMILARITY: Belongs to the annexin family.
 CC -1- SIMILARITY: Contains 4 annexin repeats.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 CC EMBL, D10883; BAA01705.1; -.
 DR PIR, JH0694; LURB11.
 DR HSSP; P13214; IANN.
 DR GO; GO:0005635; C:nuclear membrane; ISS.
 DR GO; GO:0005654; C:nucleoplasm; ISS.
 DR GO; GO:0005515; F:protein binding; ISS.
 DR InterPro; IPR001464; Annexin.
 DR Pfam; PF00191; annexin; 4.
 DR PRINTS; PR00196; ANNEXIN.
 DR ProDom; PD000143; Annexin; 4.
 DR SMART; SM00335; ANX; 4.
 DR PROSITE; PS00223; ANNEXIN; 4.
 KW Annexin, Calcium/phospholipid-binding; Repeat.
 FT REPEAT 207 267 ANNEXIN 1.
 FT REPEAT 279 339 ANNEXIN 2.
 FT REPEAT 363 423 ANNEXIN 3.
 FT REPEAT 438 498 ANNEXIN 4.
 SQ SEQUENCE 503 AA; 54034 MW; 44C15F290770AC9F CRC64;

Alignment Scores:

Pred. No.:	1.77	Length:	503
Score:	110.50	Matches:	77
Percent Similarity:	36.41%	Conservative:	53
Best Local Similarity:	21.57%	Mismatches:	123
Query Match:	3.54%	Indels:	104
DB:	1	Gaps:	17

US-10-676-079-3 (1-1721) x ANXB_RABIT (1-503)

```

QY 110 GCTCTGGGGCGCGTGGTCCCTCTCCCTCTGCGCCCTGCCCCGACC 157
   |||||
DB 76 AIAProGlyGlyGlyTyrProProValProProGlyGlyPheGlyGlnProProThr 95
   |||||
QY 158 TCGCGAAGCAGCAGCAGCTGTGGA-----CCTGGA----- 187
   |||||
DB 96 GlnProSerValProProTyrGlyValTyrProProProGlyGlyAsnProProSergly 115
   |||||
QY 188 -----CTTCTTACCAGAGCGCGTGCAGCCTGTGAGAGCCCTGCTCTGTC 235
   |||||
DB 116 ValProSerTyrProProPheProGlyAlaProValProGlyGlnProProProPro--- 134
   |||||
QY 236 CGTCAACATTGACGCCAACCTGGCCAGCGGCGGGTTCCTCATCTCTGGGTTCTCC 295
   |||||
DB 135 -----ProGlyHisGlnProProGlyProTyr---ProGlyGlnLeu 147
   |||||
QY 296 AAAGCTTGCTACTGCGCCAGAGCGTCTGCTCTGCTGCTGAGAGTTGGTGACACCA 355
   |||||
DB 148 ProValThrTyrProGlyGln-----SerProValProProPro-----GlyGln 162
   |||||
QY 356 GACAGACTTCTAATTTTGATCCCAAGAAGAAATCACTTTGAAGAGAAAGTTACTG 415
   |||||
DB 163 GlnProMetProSerTyr----- 168
   |||||
QY 416 GCAATCTCAAGTCAACGAGATATTGCAAAATATGATCATCTCTGATGTGAGAGA 475
   |||||

```

```

DB 169 -----ProGlyTyr----- 171
QY 476 GAAGTTAGGTTGGAATGGCCCTACACAGAGCA---ATTGCTACTCGAAGAACTACCA 532
   |||||
DB 172 -----ProGlySerGlyThrValThrProAlaValProPro 183
   |||||
QY 533 GAAAAGTTCAAGAACACAGCACTTCAAGAAAGCTGTAGATGTCTATCACTTTTGC 592
   |||||
DB 184 -ValGlnPheGlyAsnAlaGly-----ThrIleTh 193
   |||||
QY 593 AAAGCTGTCAGAGCTGAGCTTGATC----- 617
   |||||
DB 193 rAspAlaSerGlyPheAspProLeuArgAspAlaGluValLeuArgLysAlaMetLysG1 213
   |||||
QY 618 -TTTGCGCTTAAT-----GGTATTATTAAGAACGACAAATTGGCAGTGGAACAGTTTCA 670
   |||||
DB 213 yPheGlyThrAspGlnAlaIleIleAspCysLeuGlySerArgSerAsnLysGlnAr 233
   |||||
QY 671 TGCTCAGTTGCTCTCGAGTACTGCTCTTCCAGAGGGGTATACATTTCTTGGAACTAG 730
   |||||
DB 233 GlnGlnIleLeuLeuSerPheLysThrAlaTyrGlyAspLeuIleLysAspLeuLys 253
   |||||
QY 731 CAATGAACCTTAACAGTTCTTCTTGAAGAGCTGATATTTCATCAATGGCTGCACTTAA 790
   |||||
DB 253 sSerGlyLeuSerGlyAsnPheGlyLysThrIleLeuAlaLeuMetLysThrProIleLe 273
   |||||
QY 791 AGAAGATTATATTCATATTCATTAACCTTCAAGAAAGTCCACCTTCAAAATATGCA----- 845
   |||||
DB 273 uPheAspAlaTyrGlnIleLysGluAlaIleLeuGlyValaGlyThrAspGluAlaCysLe 293
   |||||
QY 846 -----AAAGCTATAGTCTCGATGTTGTGACGCTTCAAGAAAGAGCGGTAGATGCTGA 901
   |||||
DB 293 uIleGluIleLeuAlaSerArgSerArgGlnHisIleArgIleuAsnLysValaTyrLys 313
   |||||
QY 902 GAGC---TTCTGAAGCGTGTGAGAAAGTATGATTCAGTTACATGACATCACTAC-- 956
   |||||
DB 313 sThrGluPheLysGlyThrLeuGluGluAlaIleArgSerAspThrSerglyHisPheG1 333
   |||||
QY 957 -----TATTGATGAGACGAGCTGCTACACGAGGAATTTTCAACCTGATGATNT 1009
   |||||
DB 333 nArgLeuLeuIleSerLeuSerGlnGlyAsnArgAspLysThrAsnValAsp--We 352
   |||||
QY 1010 GGACATTTTATTTTCATCTGTGCAAAAGTTTCCAGCGTGTGTGAGAGC 1058
   |||||
DB 352 tSerLeuValGlnArgAspValGlnGluLeuTyrAlaIleGlyGluAsn 368
   |||||

```

RESULT 10

```

THIC_AGR75 STANDARD; PRT; 607 AA.
ID THIC_AGR75
AC 08UCG9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Thiamine biosynthesis protein thic.
GN THIC OR ATU2569 OR AGR C 4656.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Serubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D., St.,
RA Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavlin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krepan W., Dotan M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
"the genome of the natural genetic engineer Agrobacterium tumefaciens

```


Db	2377	Gin1leLeuLeuSerPheIyThAlaIyrcIyIySAspLeuIleIySAspLeuIySe	256
Qy	632	CTTATTA 638	
Db	256	rgIuleu 258	
RESULT 12			
ID	SPD2_NEPCL	STANDARD;	PRT; 627 AA.
AC	P46804;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Spidroin 2 (Dragline silk fibroin 2) (Fragment).		
OC	Nephila clavipes (Orb spider).		
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;		
OC	Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.		
OK	NCBI_TaxId6915;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=92406876; PubMed=1527052;		
RA	Himan M.B., Lewis R.V.;		
RT	"Isolation of a clone encoding a second dragline silk fibroin.		
RL	Nephila clavipes dragline silk is a two-protein fiber.";		
CC	J. Biol. Chem. 267:19320-19324(1992).		
CC	-1- FUNCTION: Spiders major ampullate silk possesses unique		
CC	characteristics of strength and elasticity. Fibroin consists of		
CC	pseudorectalline regions of antiparallel beta-sheet interspersed		
CC	with elastic amorphous segments.		
CC	-1- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 1, OF THE DRAGLINE SILK.		
CC	-1- SUBCELLULAR LOCATION: Extracellular.		
CC	-1- DOMAIN: Highly repetitive protein characterized by regions of		
CC	polyalanine and glycine-rich repeating units.		
CC	-1- SIMILARITY: Belongs to the silk fibroin family.		
CC	-1- DATABASE: NAME=Protein Spotlight;		
CC	NOTE=Issue 24 of July 2002;		
CC	WWW="http://www.expasy.org/spotlight/articles/sp11024.html".		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation.		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sdb.ch/announce		
CC	or send an email to license@isb-sdb.ch).		
CC	-----		
CC	EMBL; M92913; AAA29381.1; -.		
DR	PIR; A44112; A44112.		
DR	SILK; Repeat.		
FT	NON TER	1	1
FT	DOMAIN	1	530
FT	REPEAT	1	36
FT	REPEAT	37	79
FT	REPEAT	80	121
FT	REPEAT	122	172
FT	REPEAT	173	213
FT	REPEAT	214	252
FT	REPEAT	253	283
FT	REPEAT	284	317
FT	REPEAT	318	359
FT	REPEAT	360	391
FT	REPEAT	392	428
FT	REPEAT	429	464
FT	REPEAT	465	488
FT	REPEAT	489	515
FT	REPEAT	516	530
SO	SEQUENCE	627 AA;	54184 MW; CB9B63779B2C594B CRC64;
Alignment Scores:			
Pred. No.:	2.44	Length:	627
Score:	109.00	Matches:	38
Percent Similarity:	38.02%	Conservative:	8
Best Local Similarity:	31.40%	Mismatches:	37

Query Match:		3.61%	Indels:	38
DB:	1	Gaps:	4	
US-10-676-079-3 (1-1721) x SPD2_NEPCL (1-627)				
Oy	300	GCTTTGAGAACCCAGCAGATGAGAAACCGGGTCCGTGGCCAGGTGGCTCATGCG	241	
Db	315	AlAlAlAlAAGlYProglYglnGlnInlYleuGlYrGlYProglY-----	330	
Oy	240	TGACGAGACAGAAACGAGGGC---TCACCAGGTGCAGCGGTCTGGGTGAAGAATCCA	184	
Db	331	--GlnGlnGlYProglYgLYrGlYProglYglnGlnInlYProglYgLYrGlYPro	349	
Oy	183	GGTCCACGACGCTCTGTGCTTGCGAGAGTCCGGACGGGCCGACGGAGAGG----	130	
Db	350	GlYSerAlSerAl	369	
Oy	129	-----GACCCAGCGGCCCCAGGACAGCAGCATCAGCGGCGCGCACGG	85	
Db	370	GLYProglYglnGlnInlYProserGlYProglYSerAlSerAlAlAlAlAlAlAlAl	389	
Oy	84	CAGGCT-----	79	
Db	390	AlAlAlAgLYProglYgLYrGlYrGlYProglYglnGlnInlYProglYgLYrYAlAProglY	409	
Oy	78	-----TCAGCGCAGCAGCAGCATCTGGGCTCACCTGGCTGCTCCC	40	
Db	410	GlnGlnGlYProserGlYProglYSerAlSerAlAlAlAlAlAlAlAlAlAlAlAgly	429	
Oy	39	CCC 37		
Db	430	Pro 430		
RESULT 13				
RP3A_HUMAN	ID	Q9Y2J0, Q96AE0,	STANDARD;	PRT; 694 AA.
AC	Q9Y2J0,	Q96AE0,		
DT	16-OCT-2001	(Rel. 40, Created)		
DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	Rabphilin-3A (Exophilin 1).			
GN	REP3A OR KIA0985.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=992486063; PubMed=10231032;			
RA	Negase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.,			
RT	"Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro."			
RL	DNA Res. 6:63-70(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buéow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heich F., Datchenko L., Martulina K., Farmer A.A., Rubin G.W., Hong L., Stapleton M., Soares M.B., Bonald M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Uesdi T.B., Toshimaki S., Carninci P., Pirange C., Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huylk S.W., Villalón D.K., Murry D.M., Sodergren E.J., Lu X., Gibbs K.A., Fahey Y., Helton E., Ketteman M., Madan A.C., Rodríguez S., Sanchez A., Whiting M., Madan A.C., Shevchenko Y., Bouffard G.G.,			

Db	563	niglmglyleuleuilevalglylleilleargyvvalnlsyeualdaala-----	579
Qy	1238	TTTGGATGATAAACAAC-----TTGCAGCCTTACTGTATATTATGGCATCTCTTGTT	1291
Db	580	---MetserpalaaenlglyTYrSerSarpProcheVallyleutRpleu-----	594
Qy	1292	CAGAAMAATTGTGGCGCACCAAGGTGTAAATGGCAAGCGTCGAAGGTTCAAAGAGAGANA	1351
Db	595	-LyvProaerMetgilylYlvYala-----LynHiliLy 605	
Qy	1352	GCTTGCAGTAATACCTTCAATGTGCACAAACACTGCACATCCAGGTATTAAGAA	1403
Db	605	sThrcIntlelyvs-----LySlvethrlneuaSnPrrogiunPhaeangiu 619	
<hr/>			
RESULT 14			
ANXB_MOUSE			
ID	_ANXB_MOUSE	STANDARD;	PRT; 503 AA.
AC	p97384,		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Annexin A11 (Annexin XI) (calycycin-associated annexin 50) (CAP-50).		
GN	ANNX11 OR ANN11.		
OC	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxId=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RP	MEDLINE=97092887; PubMed=8938449;		
RA	Fernandez M.P., Jenkins N.A., Gilbert D.J., Copeland N.G.,		
RA	Morgan R.O.;		
RT	"Sequence and chromosomal localization of mouse annexin XI";		
RL	Genomics 37:366-374(1996).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=129/SvJ;		
RX	MEDLINE=20469408; PubMed=11013079;		
RA	Bances P., Fernandez M.R., Rodriguez-Garcia M.I., Morgan R.O.,		
RA	Fernandez M.P.;		
RT	"Annexin A11 (Annexin I) gene structure and the progenitor of paralogous		
RL	annexins and source of orthologous cDNA isoforms";		
RL	Genomics 69:95-103(2000).		
CC	-I-*FUNCTION: Binds specifically to calycycin in a calcium-dependent		
CC	manner.		
CC	-I- DOMAIN: A pair of annexin repeats may form one binding site for		
CC	calcium and phospholipid.		
CC	-I- SIMILARITY: Belongs to the annexin family.		
CC	-I- SIMILARITY: Contains 4 annexin repeats.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on ways		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to licenses@isb-sib.ch).		
CC	-----		
DR	EMBL; U65866; AAB42012.1; -		
DR	EMBL; AJ289760; CAB94770.1; -		
DR	EMBL; AJ289761; CAB94770.1; JOINED.		
DR	EMBL; AJ289762; CAB94770.1; JOINED.		
DR	EMBL; AJ289763; CAB94770.1; JOINED.		
DR	EMBL; AJ289764; CAB94770.1; JOINED.		
DR	EMBL; AJ289765; CAB94770.1; JOINED.		
DR	EMBL; AJ289766; CAB94770.1; JOINED.		
DR	EMBL; AJ289767; CAB94770.1; JOINED.		
DR	EMBL; AJ289768; CAB94770.1; JOINED.		
DR	EMBL; AJ289769; CAB94770.1; JOINED.		
DR	HSSP; P13214; 1ANN.		
DR	SWISS-2DPAGE; P97384; MOUSE.		
DR	MCD; MG1:108481; Annx11.		
DR	GO; GO:0005635; C:nuclear membrane, ISS.		

DR	GO: 0005654; C:nucleoplasm; ISS.
DR	GO: 0005515; R:protein binding; ISS.
DR	InterPro: IP001464; Annexin.
DR	Pfam: PF00151; annexin; 4.
DR	PRINTS: PR00196; ANNEXIN.
DR	ProDom: PD000143; Annexin; 4.
DR	SMART: SM00335; ANX; 4.
DR	PROSITE: PS00223; ANNEXIN; 4.
KW	Annexin; Calcium/phospholipid-binding; Repeat.
FT	REPEAT 207 267 ANNEXIN 1.
FT	REPEAT 279 339 ANNEXIN 2.
FT	REPEAT 363 423 ANNEXIN 3.
FT	REPEAT 438 498 ANNEXIN 4.
SEQ	SEQUENCE 503 AA; 5411 MW; 424B1345E0F4EC8A CnC64;
Alignment Scores:	
Pred. No.:	2.72
Score:	108.00
Percent Similarity:	37.85%
Best Local Similarity:	20.06%
Query Match:	3.46%
DB:	1
Length:	
Matches:	503
Conservative:	71
Mismatches:	63
Indels:	125
Gaps:	96
	12

[illegible]

QY 740 TAACAGTTCCTTAAGAGCTGATATTTTCATCAATGGGTGCGAGTTAGAGAGATTA 799
 Db 256 userglyanphegyluvsthrilieleuallaleuetythrprovalleupheaspya 276
 QY 800 TATTCAATTCGCTAAATCTTAAAGAACTCCACTTTCAAAAATGCA-----AAACT 850
 Db 276 ltyrluilelysglualalelysglyalaglythraepglualacylsleuileglul 296
 QY 851 CTATGTCCTGATGTTGGTCGAGCTCGAAGAAAGCGGTAGATAGATGCTAAGAGC---TT 907
 Db 256 ephelaaseraragseraengluhisilexrgluuuseratgalaityrlysthrglu 316
 QY 908 CCTGAGGCTGTGTGAGAGATGATTGATTTCAGTTCATGCGATCACTAC-----TA 958
 Db 316 eglulnysrthrleugluualaleargseraphrthrsergylhispheglaryleu 336
 QY 959 TTTGATGACGAGCTGCTACACGAGAGATTTTCAAACTGATGATGACATTTT 1018
 Db 336 uleserleuserglinglyasnaragapgluserthravenalasp---Metserleuva 355
 QY 1019 TATTTCATCTGTCAAAAAGTTTCCAGGTGTGAGAGC 1058
 Db 355 lglmaragapvalingluileuylalaglygluasn 368

RESULT 15
 SYM_CLOTE ID SYM_CLOTE STANDARD; PRT; 641 AA.
 AC 0895D9;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Methionyl-tRNA synthetase (EC 6.1.1.10) (Methionine--tRNA ligase) (MeRS).
 GN MERS OR CTC00243.
 OS Clostridium tectani.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxId=1513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Massachusetts / E88;
 RX MEDLINE=22457253; PubMed=12552129;
 RA Brueggemann H., Baesner S., Fricke W.P., Wierer A., Liesegang H.,
 RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
 RA Gottschalk G.;
 RT "The genome sequence of Clostridium tectani, the causative agent of
 RT tetanus disease.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
 CC -!- FUNCTION: Is required not only for elongation of protein synthesis
 CC but also for the initiation of all mRNA translation through
 CC initiator tRNA (fMet) aminoacylation (By similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA(Met) = AMP +
 CC diphosphate + L-methionyl-tRNA(Met).
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -!- SUBUNIT: Homodimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC Meig subfamily 2A.
 CC -!- SIMILARITY: Contains 1 tRNA-binding domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; AE015936; AAO34890.1; -
 DR HAMAP; MF_01228; -; 1.
 DR InterPro; IPR008994; Nucleic_acid_OB.
 DR InterPro; IPR002300; tRNA-synt_1a.
 DR InterPro; IPR001412; tRNA-synt_1.

DR InterPro; IPR002304; tRNA-synt met.
 DR InterPro; IPR002547; tRNA bind_
 DR Pfam; PF00133; tRNA-synt_I; 1.
 DR Pfam; PF01586; tRNA_bind_1.
 DR PRINTS; PR01041; TRNASYNTHET.
 DR PROSITE; PS00178; AA tRNA_LIGASE_I; 1.
 DR PROSITE; PS50886; TRSD; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW tRNA-binding; Metal-binding; Zinc; Complete proteome.
 FT SITE 13 23
 FT SITE 298 302 "HIGH" REGION.
 FT SITE 539 641 "KMSKS" REGION.
 FT DOMAIN 539 641 tRNA-BINDING.
 FT METAL 128 128 ZINC (BY SIMILARITY).
 FT METAL 131 131 ZINC (BY SIMILARITY).
 FT METAL 145 145 ZINC (BY SIMILARITY).
 FT METAL 148 148 ZINC (BY SIMILARITY).
 FT BINDING 301 301 ATP (BY SIMILARITY).
 SQ SEQUENCE 641 AA; 73211 MW; B58300AA16B20AC4 CRC64;

Alignment Scores:

Pred. No.: 2.92 Length: 641
 Score: 108.00 Matches: 76
 Percent Similarity: 33.33% Conservative: 56
 Best Local Similarity: 19.19% Mismatches: 146
 Query Match: 3.46% Indels: 118
 Ds: 1 Gaps: 16

US-10-676-079-3 (1-1721) x SYM_CLOTE (1-641)

QY 549 AGCACTACTCAAGAGAGCTGTAGATGCTATACACTTTGCAAACTGCTAGAGACTG 608
 Db 23 aenthrlyrthrthrvalalalaspalaleuualargphelyargleuthrcllyhis 42
 QY 609 GACTTGATCTTTGGCTTAATGCCATTATTAAGACAGCAAGTTTCACTGCAACGTTCT 668
 Db 43 aspvallleuphe-----Leuthrcllythraepglu 52
 QY 669 AATGCTGATGCTGCTGAGACTACTGCTCTTCCAGAGGGTATAACATTCTTGAGACTA 728
 Db 53 hisgllylnlyslleaglaryvalalagluclulysgllyleuylporysglyluyrleu 72
 QY 729 GGCATGCAACTCAACATGCTTCTTAG-----AAGCTGATATT 767
 Db 73 aspaenmetvalaspserillelysgluuethrlyssermetaenlleserlyrbspyls 92
 QY 768 TTCATCAATGGGTGCGAGTTCAGAGAGATTAATTTCAATTGCAATTAACCTTAAGA 824
 Db 93 pheileargthrthr-----Aspapyrthhisilegluservalglulysilephe 109
 QY 825 AAGTCCACCTTCAAAAATGCAAACTCTATGCT----- 857
 Db 110 lyslyslleuyltyrgluinglyaspillelyrlysglyluyrlygluluyrtyrcysthr 129
 QY 858 -----CTGATGTTGAT 869
 Db 130 ProCysgluserphetrthrclusercInleuaspaphisbenCysProaspCysely 149
 QY 870 CAGCTCGAAGAAAGCGCTTAAGATGCTGAAGACTTC-----CTGAGCTGTGTGGA 923
 Db 150 ArgProvalgluluyrthrlysgluclulalatyrrhepelysmetserlytyrvalasp 169
 QY 924 GAAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 983
 Db 170 ArgleuilelyrtyrileglulugluasnProhisPheileglInProgluserarglyshen 189
 QY 984 GAA-----GATTTCTTAACCTGATGATGATGATGATGATGATGATGATGATGATGAT 1034
 Db 190 GlumetleuenaenpheleuylserProcllyleuInaspheCysIleser----- 206
 QY 1035 AAAGTTTCCAGGTGTGTTGAGAGCACAGGCTGGCAAGAAAGTCTGTTAGAGAAACA 1094
 Db 206 ----- 206

```
QY 1095 AGCTTCGATATGAGCGGAGCCCTTGTCTATCCGACACTTTGCACTGGCTTATG 1154
Db : : : : :
QY 207 ArgThrSerPheAtpGlyIleProValSerPheAspAsnLysHisValIleTyrVal 1226
Db : : : : :
QY 1155 TGGCTGGATTAATG---GGCTGTGAGCCGGAATGGGAATAGAAAGTGATGAGGCA 1211
Db : : : : :
QY 227 TrpIleAspAlaLeuSerAsnTyrIleThrAlaLeuGlyTyrAsnSerAspAsnGlnGlu 246
Db : : : : :
QY 1212 GATATTC---TTTGAGAGCAGAAACTATACATTAGTGATGAAAC-----TTC 1256
Db : : : : :
QY 247 LeuLeuGlnLysPheTyrProAlaAsnValHisLeuValGlyLysAspIleLeuArgPhe 266
QY 1257 GATCCTTACCTGATTTATGGCTATCTCTGTTC-----AAG 1295
Db : : : : :
QY 267 HisThrIle---TyrTrpProIleMetLeuMetAlaLeuGlyIleGlnLeuProLys 284
QY 1296 AAATTGGTGGGACCAAGAGTGTATATGCAAGCGTGCAAGGTTCCAAGAGAGAGACTT 1355
Db : : : : :
QY 285 GlnValPheGlyHisGlyTrpLeuLeu-----ValAspGlyGlyLysMetSerLysSer 302
QY 1356 CGAGTATACCTTCATTGACACAAACATGACATCCAAAGTATTAAGAGAGATTAACT 1415
Db : : : : :
QY 303 Lys----- 303
QY 1416 CTGTATGCCATAAACCCTCATTAAGCTCACCAAGTACTTGGGTTACCTTCTTCT 1475
Db : : : : :
QY 304 -----GlyAsnValValAspProValValLeuValAspHisPheGly 317
QY 1476 AACAGCAGATGGATTAATACCTTCTTAAGA-----CCTTGGGAGCTCATGGATTACTT 1529
Db : : : : :
QY 318 GluAspThrValArgTyrTyrLeuLeuArgGlnIleProPheGlySerAspGlyLeuPhe 337
QY 1530 TCCAAATCTGTCCAACTCAATGCTTAACCTTAAGATGGTGATGAT 1577
Db : : : : :
QY 338 AsnAsnGlnLeuPheIleLysLysIleAsnSerAspLeuAlaAsnAsp 353
```

Search completed: August 28, 2004, 07:44:22
Job time : 61.5 secs